



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/12, 15/63, 5/16, C12P 21/08, C07K 16/18, 16/40</b>		<b>A2</b>	(11) International Publication Number: <b>WO 97/12039</b>
			(43) International Publication Date: <b>3 April 1997 (03.04.97)</b>
(21) International Application Number: <b>PCT/CA96/00655</b> (22) International Filing Date: <b>27 September 1996 (27.09.96)</b> (30) Priority Data: 60/006,063           27 September 1995 (27.09.95)   US 60/007,788           30 November 1995 (30.11.95)   US 60/015,217           9 April 1996 (09.04.96)       US 08/664,962           14 June 1996 (14.06.96)       US (60) Parent Application or Grant (63) Related by Continuation US                               08/664,962 (CON) Filed on                       14 June 1996 (14.06.96) (71)(72) Applicant and Inventor: <b>KRYSTAL, Gerald [CA/CA]; 601 West 10th Avenue, Vancouver, British Columbia V5Z 1L3 (CA).</b> (74) Agent: <b>BERESKIN &amp; PARR; 40th floor, 40 King Street West, Toronto, Ontario M5H 3Y2 (CA).</b>		(81) Designated States: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</b>  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>	
(54) Title: <b>SH2-CONTAINING INOSITOL-PHOSPHATASE</b>			
(57) Abstract			
<p>Novel SH2-containing inositol-phosphatase which has a src homology 2 (SH2) domain and exhibits phospholns-5-ptase activity, and nucleic acid molecules encoding the novel protein are disclosed. The invention also relates to methods for identifying substances which affect the binding of the protein to Shc and/or its phospholns-5-ptase activity and methods for screening for agonists or antagonists of the binding of the protein and Shc.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

**Title: SH2-CONTAINING INOSITOL-PHOSPHATASE****FIELD OF THE INVENTION**

The invention relates to a novel SH2-containing inositol-phosphatase, truncations, analogs, homologs and isoforms thereof; nucleic acid molecules encoding the protein and truncations, analogs, and homologs of the protein; and, uses of the protein and nucleic acid molecules.

**BACKGROUND OF THE INVENTION**

Many growth factors regulate the proliferative, differentiative and metabolic activities of their target cells by binding to, and activating cell surface receptors that have tyrosine kinase activity (Cantley, L.C., et al. 1991, Cell 64:281-302; and Ullrich, A., and J. Schlessinger. 1990, Cell 61:203-212). The activated receptors become tyrosine phosphorylated through intermolecular autophosphorylation events, and then stimulate intracellular signalling pathways by binding to, and phosphorylating cytoplasmic signalling proteins (Cantley, L.C., et al. 1991, Cell 64:281-302; and, Ullrich, A., and J. Schlessinger, 1990, Cell 61:203-212). Many cytoplasmic signalling proteins share a common structural motif, known as the src homology 2 (SH2) domain, that mediates their association with specific phosphotyrosine-containing sites on activated receptors (Heldin, C.H. 1991, Trends Biochem. Sci. 16:450-452; Koch, C.A., et al., 1991, Science 252:669-674; Margolis, B. 1992, Cell Growth Differ. 3:73-80; McGlade, C.J., et al, 1992, Mol. Cell. Biol. 12: 991-997; Moran, M.F., et al., 1990, Proc. Natl. Acad. Sci. USA 87:8622-8626; and Reedijk, M., et al., 1992, EMBO J. 11:1365-1372).

Two SH2-containing proteins, Grb2 and Shc, have been implicated in the Ras signalling pathway (Lowenstein, E.J., et al., 1992, Cell 70:431-442, and, Pelicci, G., et al., 1992, Cell 70 93-104.). Grb2 and Shc act upstream of Ras and bind directly to activated receptors (Buday, L., and J. Downward, 1993, Cell 73:611-620; Matuoka, K. et al., 1993, EMBO J. 12:3467-3473, Oakley, B.R. et al., 1980, Anal. Biochem. 105:361-363., Reedijk, M., et al., 1992, EMBO J. 11:1365-1372; Rozakis-Adcock, M., et al., 1992 Nature 360: 689-692; and, Songyang, Z., et al., 1993, Cell 72:767-778), or to designated SH2 docking proteins, such as the insulin receptor substrate 1 (IRS-1), which is tyrosine phosphorylated in response to insulin (Baltensperger, K., et al., Science 260:1950-1952; Pelicci, G., et al., 1992, Cell 70:93-104; Skolnik, E.Y., 1993, EMBO J. 12:1929-1936; Skolnik, E.Y., et al., 1993, Science 260:1953-1955; and Suen, K-L., et al., 1993 Mol. Cell. Biol. 13: 5500-5512).

Grb2 is a 25 kDa adapter protein with two SH3 domains flanking one SH2 domain. It has been shown in fibroblasts to shuttle its constitutively bound Ras guanine nucleotide exchange factor, Sos1, to activated receptors (or to IRS-1 (Skolnik, E.Y., 1993, EMBO J. 12:1929-1936; and Skolnik, E.Y., et al., 1993, Science 260:1953-1955), (Baltensperger, K., et al., Science 260:1950-1952; Buday, L., and J. Downward, 1993, Cell 73:611-620; Egan, S.E. et al., 1993, Nature (London) 367:87-90; Gale, N.W., et al., 1993, Nature (London) 363:88-92; Li, N., et al.,

1993, *Nature* (London) 363:85-88; Olivier, J.P. et al., 1993, *Cell* 73:179-191; and Rozakis-Adcock, M., et al., 1993 *Nature* (London) 363:83-85). Binding of the SH2 domain of Grb2 to tyrosine phosphorylated proteins activates Sos1 which then catalyzes the activation of Ras by exchanging GDP for GTP (Buday, L., and J. Downward. 1993. *Cell* 73:611-620 12,,20; Egan, S.E. Et al, 1993, *Nature* 363:45-51; Gale, N.W et al., 1993 *Nature* 363:88-92; Li, N., et al., 1993 *Nature* 363:85-88).

Shc is also an adapter protein that is widely expressed in all tissues. The protein contains an N-terminal phosphotyrosine binding (PTB) domain (Kavanaugh, V.M. Et al., 1995 *Science*, 268:1177-1179; Craparo, A., et al., 1995, *J. Biol. Chem.* 270:15639-15643; van der Geer, P., & Pawson, T., 1995, *TIBS* 20:277-280; Batzer, A.G., et al., *Mol. Cell. Biol.* 1995, 15:4403-4409; and Trub, T., et al., 1995, *J. Biol. Chem.* 270:18205-18208) and a C-terminal SH2 domain (Pelicci, G., et al., 1992. *Cell* 70:93-104) and can associate, in its tyrosine phosphorylated form, with Grb2-Sos1 complexes and may increase Grb2-Sos1 interactions following growth factor stimulation (Egan, S.E. Et al, 1993, *Nature* 363:45-51; Rozakis-Adcock, M., et al., 1992, *Nature* 360:689-692; and Ravichandran, K.S., 1995, *Mol. Cell. Biol.* 15:593-600). Shc appears to function as a bridge between Grb2-Sos1 complexes and tyrosine kinases where the latter are incapable, for lack of an appropriate consensus sequence, of binding Grb2-Sos1 directly (Egan, S.E. Et al, 1993, *Nature* 363:45-51).

Preliminary evidence suggests that Shc and Grb2 may be used by members of the hemopoietin receptor superfamily (Cutler, R.L., et al., 1993, *J. Biol. Chem.* 268:21463-21465, Damen, J.E., et al., 1993, *Blood* 82:2296-2303). Although members of this family lack endogenous kinase activity, following ligand binding, they are apparently tyrosine phosphorylated by a closely associated JAK family member (Argetsinger, L.S., et al., 1993, *Cell* 74:237-244; Lutticken, C., et al., 1994, *Science* 263:89-92; Silvennoinen, O., et al., 1993, *Proc. Natl. Acad. Sci. USA* 90:8429-8433; and Witthuhn, B.A., et al., 1993, *Cell* 74:227-236). The hemopoietic growth factors, erythropoietin (Ep), interleukin-3 (IL-3) and steel factor (SF) (which utilizes a receptor with endogenous tyrosine kinase activity, i.e., c-kit, (Chabot, B., et al., 1988, *Nature* (London) 335:88-89)), have been shown to induce the tyrosine phosphorylation of Shc and its subsequent association with Grb2 (Cutler, R.L., et al., 1993, *J. Biol. Chem.* 268:21463-21465). Stimulation of members of the hemopoietin receptor superfamily has also been reported to result in the association of Shc with uncharacterized proteins with molecular masses of 130 kDa (Smit, L., et al., *J. of Biol. Chem.* 269(32):20209, 1994), 150 kDa (Lioubin, M.N., et al., *Mol. Cell. Biol.* 14(9):5682, 1994), and 145 kDa (Damen, J., et al., *Blood* 82(8):2296, 1993, and Saxton, T.M. et al., *J. Immunol.* 623, 1994).

## **SUMMARY OF THE INVENTION**

The present inventor has identified and characterized a protein that associates with Shc in response to multiple cytokines. The unique protein, herein referred to as "SH2-containing inositol-phosphatase" or "SHIP" (for SH2-containing, inositol 5-phosphatase),

contains an amino terminal src homology 2 (SH2) domain, two phosphotyrosine binding (PTB) consensus sequences, a proline rich region, and two motifs highly conserved among inositol polyphosphate-5-phosphatases (phosphoIns-5-ptases). Cell lysates immunoprecipitated with antiserum to the protein exhibit phosphoIns-5-ptase activity, in particular, both  
5 phosphatidylinositol trisphosphate (PtdIns-3,4,5-P<sub>3</sub>) and inositol tetraphosphate (Ins-1,3,4,5-P<sub>4</sub>) 5-phosphatase activity. This activity implicates SHIP in the regulation of signalling pathways that control gene expression, cell proliferation, differentiation, activation, and metabolism, in particular, the Ras and phospholipid signalling pathways. This finding permits the identification of substances which affect SHIP and which may be  
10 used in the treatment of conditions involving perturbation of signalling pathways.

The present invention therefore provides a purified and isolated nucleic acid molecule comprising a sequence encoding an SH2-containing inositol-phosphatase which has a src homology 2 (SH2) domain and exhibits phosphoIns-5-ptase activity. The SH2-containing inositol-phosphatase is further characterized by its ability to associate with Shc and by  
15 having two phosphotyrosine binding (PTB) consensus sequences, a proline rich region, and motifs highly conserved among inositol polyphosphate-5-phosphatases (phosphoIns-5-ptases).

In an embodiment of the invention, the purified and isolated nucleic acid molecule comprises (i) a nucleic acid sequence encoding an SH2-containing inositol-phosphatase having  
20 the amino acid sequence as shown in SEQ ID NO:2 or Figure 2 (A); and, (ii) nucleic acid sequences complementary to (i). In another embodiment of the invention, the purified and isolated nucleic acid molecule comprises (i) a nucleic acid sequence encoding an SH2-containing inositol-phosphatase having the amino acid sequence as shown in SEQ ID NO:8 or Figure 11; and, (ii) nucleic acid sequences complementary to (i).

25 In a preferred embodiment of the invention, the purified and isolated nucleic acid molecule comprises

- (i) a nucleic acid sequence encoding an SH2-containing inositol-phosphatase having the nucleic acid sequence as shown in SEQ ID NO:1 or Figure 3, wherein T can also be U;
- (ii) a nucleic acid sequence complementary to (i), preferably complementary to the full  
30 length nucleic acid sequence shown in SEQ ID NO: 1 or Figure 3; or
- (iii) a nucleic acid molecule differing from any of the nucleic acids of (i) and (ii) in codon sequences due to the degeneracy of the genetic code.

In another preferred embodiment of the invention, the purified and isolated nucleic acid molecule comprises

- 35 (i) a nucleic acid sequence encoding an SH2-containing inositol-phosphatase having the nucleic acid sequence as shown in SEQ ID NO:7 or Figure 10, wherein T can also be U;
- (ii) a nucleic acid sequence complementary to (i), preferably complementary to the full length nucleic acid sequence shown in SEQ ID NO: 7 or Figure 10;

(iii) a nucleic acid molecule differing from any of the nucleic acids of (i) and (ii) in codon sequences due to the degeneracy of the genetic code.

The invention also contemplates (a) a nucleic acid molecule comprising a sequence encoding a truncation of the SH2-containing inositol-phosphatase, an analog or homolog of the SH2-containing inositol-phosphatase or a truncation thereof, (herein collectively referred to as "SHIP related protein" or "SHIP related proteins"); (b) a nucleic acid molecule comprising a sequence which hybridizes under high stringency conditions to the nucleic acid encoded by a SH2-containing inositol-phosphatase having the amino acid sequence as shown in SEQ ID NO:2 or Figure 2 (A), or SEQ ID NO:8 or Figure 11, wherein T can also be U, or complementary sequences thereto, or by a SHIP related protein; and (c) a nucleic acid molecule comprising a sequence which hybridizes under high stringency conditions to the nucleic acid encoded by the SH2-containing inositol-phosphatase having the nucleic acid sequence as shown in SEQ ID NO:1 or Figure 3, or SEQ ID NO:7 or Figure 10, wherein T can also be U, or complementary sequences thereto.

The invention further contemplates a purified and isolated double stranded nucleic acid molecule containing a nucleic acid molecule of the invention, hydrogen bonded to a complementary nucleic acid base sequence.

The nucleic acid molecules of the invention may be inserted into an appropriate expression vector, i.e. a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Accordingly, recombinant expression vectors adapted for transformation of a host cell may be constructed which comprise a nucleic acid molecule of the invention and one or more transcription and translation elements operatively linked to the nucleic acid molecule.

The recombinant expression vector can be used to prepare transformed host cells expressing SH2-containing inositol-phosphatase or a SHIP related protein. Therefore, the invention further provides host cells containing a recombinant molecule of the invention. The invention also contemplates transgenic non-human mammals whose germ cells and somatic cells contain a recombinant molecule comprising a nucleic acid molecule of the invention which encodes an analog of SH2-containing inositol-phosphatase, i.e. the protein with an insertion, substitution or deletion mutation.

The invention further provides a method for preparing a novel SH2-containing inositol-phosphatase, or a SHIP related protein utilizing the purified and isolated nucleic acid molecules of the invention. In an embodiment a method for preparing an SH2-containing inositol-phosphatase or a SHIP related protein is provided comprising (a) transferring a recombinant expression vector of the invention into a host cell; (b) selecting transformed host cells from untransformed host cells; (c) culturing a selected transformed host cell under conditions which allow expression of the SH2-containing inositol-phosphatase or SHIP

related protein; and (d) isolating the SH2-containing inositol-phosphatase or SHIP related protein.

The invention further broadly contemplates a purified and isolated SH2-containing inositol-phosphatase which contains an SH2 domain and which exhibits phosphoIns-5-ptase activity. In an embodiment of the invention, a purified SH2-containing inositol-phosphatase is provided which has the amino acid sequence as shown in SEQ ID NO:2 or Figure 2 (A). In another embodiment of the invention, a purified SH2-containing inositol-phosphatase is provided which has the amino acid sequence as shown in SEQ ID NO:8 or Figure 11. The purified and isolated protein of the invention may be activated i.e. phosphorylated. The invention also includes truncations of the protein and analogs, homologs, and isoforms of the protein and truncations thereof (i.e. "SHIP related proteins").

The SH2-containing inositol-phosphatase or SHIP related proteins of the invention may be conjugated with other molecules, such as proteins to prepare fusion proteins. This may be accomplished, for example, by the synthesis of N-terminal or C-terminal fusion proteins.

The invention further contemplates antibodies having specificity against an epitope of SH2-containing inositol-phosphatase or a SHIP related protein of the invention. Antibodies may be labelled with a detectable substance and they may be used to detect the SH2-containing inositol-phosphatase or a SHIP related protein of the invention in tissues and cells.

The invention also permits the construction of nucleotide probes which are unique to the nucleic acid molecules of the invention and accordingly to SHIP or a SHIP related protein of the invention. Thus, the invention also relates to a probe comprising a sequence encoding SH2-containing inositol-phosphatase or an SHIP related protein. The probe may be labelled, for example, with a detectable substance and it may be used to select from a mixture of nucleotide sequences a nucleotide sequence coding for a protein which displays one or more of the properties of SHIP.

The invention still further provides a method for identifying a substance which is capable of binding to SHIP, or a SHIP related protein or an activated form thereof, comprising reacting SHIP, or a SHIP related protein, or an activated form thereof, with at least one substance which potentially can bind with SHIP, or a SHIP related protein or an activated form thereof, under conditions which permit the formation of complexes between the substance and SHIP or SHIP related protein or an activated form thereof, and assaying for complexes, for free substance, for non-complexed SHIP or SHIP related protein or an activated form thereof, or for activation of SHIP.

Still further, the invention provides a method for assaying a medium for the presence of an agonist or antagonist of the interaction of SHIP, or a SHIP related protein or an activated form thereof, and a substance which binds to SHIP, a SHIP related protein or an activated form thereof. In an embodiment, the method comprises providing a known concentration of

SHIP, or a SHIP related protein, with a substance which is capable of binding to SHIP, or SHIP related protein and a test substance under conditions which permit the formation of complexes between the substance and SHIP, or SHIP related protein, and assaying for complexes, for free substance, for non-complexed SHIP or SHIP related protein, or for  
5 activation of SHIP, or SHIP related protein. In a preferred embodiment of the invention, the substance is Shc or a part thereof, or an SH3-containing protein or part thereof.

Still further the invention contemplates a method for assaying for the affect of a substance on the phosphoIns-5-ptase activity of SHIP or a SHIP related protein having phosphoIns-5-ptase activity comprising reacting a substrate which is capable of being  
10 hydrolyzed by SHIP or a SHIP related protein to produce a hydrolysis product, with a test substance under conditions which permit the hydrolysis of the substrate, determining the amount of hydrolysis product, and comparing the amount of hydrolysis product obtained with the amount obtained in the absence of the substance to determine the affect of the substance on the phosphoIns-5-ptase activity of SHIP or the SHIP related protein.

15 Substances which affect SHIP or a SHIP related protein may also be identified using the methods of the invention by comparing the pattern and level of expression of SHIP or a SHIP related protein of the invention in tissues and cells in the presence, and in the absence of the substance.

The substances identified using the method of the invention may be used in the  
20 treatment of conditions involving the perturbation of signalling pathways, and in particular in the treatment of proliferative disorders. Accordingly, the substances may be formulated into pharmaceutical compositions for administration to individuals suffering from one of these conditions.

Other objects, features and advantages of the present invention will become apparent  
25 from the following detailed description. It should be understood, however, that the detailed description and the specific examples while indicating preferred embodiments of the invention are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

### 30 **DESCRIPTION OF THE DRAWINGS**

The invention will be better understood with reference to the drawings in which:

Figure 1 are immunoblots showing lysates prepared from B6SutA<sub>1</sub> cells, treated  $\pm$  IL-3, immunoprecipitated with anti-Shc, followed by protein A Sepharose (lanes 1&2) or incubated with GSH bead bound GST-N-SH3 (lanes 3&4) or GSH bead bound GST-C-SH3  
35 (lanes 5&6);

Figure 2 shows the amino acid sequence of murine SHIP (A) and a schematic diagram of the domains of the novel protein of the invention (B);

Figure 3 shows the nucleic acid sequence of murine SHIP;

Figure 4 shows immunoblots of lysates from B6SUA<sub>1</sub> cells, treated  $\pm$  IL-3, immunoprecipitated with anti-Shc (lanes 1&2), NRS (lanes 3&4) or anti-15mer (lanes 5&6) or precleared with anti-15mer and then immunoprecipitated with anti-Shc (lanes 7&8) (A); and lysates from B6SUA<sub>1</sub> cells, stimulated with IL-3, immunoprecipitated with anti-Shc (lane 1) or anti-15mer (lane 2) and bound proteins eluted with SDS-sample buffer containing N-ethylmaleimide in lieu of 2-mercaptoethanol (B);

Figure 5 shows Northern blot analysis of 2  $\mu$ g of polyA RNA from various tissues probed with a random primer-labeled PCR fragment encompassing a 1.5-kb fragment corresponding to the 3' end of the p145 cDNA (lanes 1-6, spleen, lung, liver, skeletal muscle, kidney and testes, respectively (Clontech); lane 7, separately prepared blot of bone marrow;

Figure 6 is a graph showing the results of anti-15mer, anti-Shc and NRS immunoprecipitates with B6SUA<sub>1</sub> cell lysate incubated with [<sup>3</sup>H]Ins-1,3,4,5-P<sub>4</sub> under conditions where product formation was linear with time (A); and shows immunoblots of anti-15mer, NRS and anti-Shc immunoprecipitates (as well as  $\pm$  recombinant 5-ptase II, ie. PtlI&BL (blank)) incubated with PtdIns[<sup>32</sup>P]-3,4,5-P<sub>3</sub> under conditions where product formation was linear with time and the reaction mixture chromatographed on TLC(B);

Figure 7 shows the amino acid sequence of Shc;

Figure 8 shows the nucleic acid sequence of Shc;

Figure 9 shows the amino acid and nucleic acid sequences of Grb2;

Figure 10 shows the nucleic acid sequence of human SHIP;

Figure 11 shows the amino acid sequence of human SHIP;

Figure 12 shows a comparison of the amino acid sequences of human and murine SHIP;

and

Figure 13 shows a comparison of the nucleic acid sequences of human and murine SHIP.

## **DETAILED DESCRIPTION OF THE INVENTION**

The following standard abbreviations for the amino acid residues are used throughout the specification: A, Ala - alanine; C, Cys - cysteine; D, Asp- aspartic acid; E, Glu - glutamic acid; F, Phe - phenylalanine; G, Gly - glycine; H, His - histidine; I, Ile - isoleucine; K, Lys - lysine; L, Leu - leucine; M, Met - methionine; N, Asn - asparagine; P, Pro - proline; Q, Gln - glutamine; R, Arg - arginine; S, Ser - serine; T, Thr - threonine; V, Val - valine; W, Trp - tryptophan; Y, Tyr - tyrosine; and p.Y., P.Tyr - phosphotyrosine.

### **I. Nucleic Acid Molecules of the Invention**

As hereinbefore mentioned, the invention provides an isolated and purified nucleic acid molecule having a sequence encoding an SH2-containing inositol-phosphatase (SHIP) which contains an SH2 domain and exhibits phosphoIns-5-ptase activity. The term "isolated and purified" refers to a nucleic acid substantially free of cellular material or culture medium when produced by recombinant DNA techniques, or chemical precursors, or other chemicals when chemically synthesized. An "isolated and purified" nucleic acid is also substantially

free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) from which the nucleic acid is derived. The term "nucleic acid" is intended to include DNA and RNA and can be either double stranded or single stranded.

The murine SHIP coding region was cloned by purifying the protein based on Grb2-C-SH3 affinity chromatography. An unambiguous sequence obtained from the purified protein, VPAEGVSSLNEMINP, was used to construct a degenerate oligonucleotide probe. The full length cDNA was cloned using a PCR based strategy and a B6SUtA<sub>1</sub> cDNA library as more particularly described in the Example herein. The nucleic acid sequence of murine SHIP is shown in Figure 3 or in SEQ. I.D. NO. 1. The underlined ATG is the likely start site (starting at nucleic acid 139). However, the predicted protein sequence shown in Figure 2 (A) (SEQ.ID.NO. 2) is from an in frame ATG starting slightly upstream at nucleotide 130. The nucleotides from approximately 151 to 444 code for the SH2 domain; the nucleotides from 1886 to 1934, and 2144 to 2167 code for 5-phosphatase motifs; the nucleotides from 1783 to 2130 code for the 5-ptase domain; nucleotides 2866-2880 and 3175 to 3189 code for the PTB domain target sequences, INPNY and ENPLY; and, the nucleotides 3013 to 3580 code for the proline-rich domain.

The nucleic acid sequence of human SHIP is shown in Figure 10 and and Figure 13 (or in SEQ.ID.NO. 7). The human SHIP gene was mapped to chromosome 2 at the junction between q36 and q37. The nucleotides from approximately 141 to 434 in Figure 10 (SEQ.ID.NO. 7) code for the SH2 domain; the nucleotides from 1876 to 1924 and 2134 to 2157 in Figure 10 code for 5-phosphatase motifs; the nucleotides from 1773 to 2120 in Figure 10 code for the 5-ptase domain; nucleotides 2856 to 2870 and 3177 to 3191 in Figure 10 code for the PTB domain target sequences, INPNY and ENPLY; and the nucleotides 3009 to 3564 in Figure 10 code for the proline-rich domain. Figure 13 shows a comparison of the nucleic acid sequences encoding human SHIP and murine SHIP. The nucleic acid sequences encoding human and murine SHIP are 81.6% identical.

The invention includes nucleic acids having substantial homology or identity with the nucleic acid sequences encoding human and murine SHIP. Homology or identity refers to sequence similarity between the nucleic acid sequences and it may be determined by comparing a position in each sequence which is aligned for purposes of comparison. When a position in the compared sequence is occupied by the same nucleotide base, then the molecules are identical or homologous at that position.

It will be appreciated that the invention includes nucleic acid molecules encoding truncations of SHIP, and analogs and homologs of SHIP and truncations thereof (i.e., SHIP related proteins), as described herein. It will further be appreciated that variant forms of the nucleic acid molecules of the invention which arise by alternative splicing of an mRNA corresponding to a cDNA of the invention are encompassed by the invention.

Another aspect of the invention provides a nucleic acid molecule which hybridizes under high stringency conditions to a nucleic acid molecule which comprises a sequence which encodes SHIP having the amino acid sequence shown in Figure 2 (A) or SEQ ID NO:2, or Figure

11 or SEQ ID NO:8, or to a SHIP related protein, and preferably having the activity of SHIP. Appropriate stringency conditions which promote DNA hybridization are known to those skilled in the art, or can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, 6.0 x sodium chloride/sodium citrate (SSC) at about  
5 45°C, followed by a wash of 2.0 x SSC at 50°C may be employed. The stringency may be selected based on the conditions used in the wash step. By way of example, the salt concentration in the wash step can be selected from a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be at high stringency conditions, at about 65°C.

10 Isolated and purified nucleic acid molecules encoding a protein having the activity of SHIP as described herein, and having a sequence which differs from the nucleic acid sequence shown in SEQ ID NO:1 or Figure 3, or SEQ ID NO:7 or Figure 10, due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode functionally equivalent proteins (e.g., a protein having SH2-containing inositol-phosphatase activity) but  
15 differ in sequence from the sequence of SEQ ID NO:1 or Figure 3, or SEQ ID NO:7 or Figure 10, due to degeneracy in the genetic code.

In addition, DNA sequence polymorphisms within the nucleotide sequence of SHIP (especially those within the third base of a codon) may result in "silent" mutations in the DNA which do not affect the amino acid encoded. However, DNA sequence polymorphisms  
20 may lead to changes in the amino acid sequences of SHIP within a population. It will be appreciated by one skilled in the art that these variations in one or more nucleotides (up to about 3-4% of the nucleotides) of the nucleic acids encoding proteins having the activity of SHIP may exist among individuals within a population due to natural allelic variation. Any and all such nucleotide variations and resulting amino acid polymorphisms are within the  
25 scope of the invention.

An isolated and purified nucleic acid molecule of the invention which comprises DNA can be isolated by preparing a labelled nucleic acid probe based on all or part of the nucleic acid sequence shown in SEQ ID NO: 1 or Figure 3, (for example, nucleotides 2830 to 2874 encoding VPAEGVSSLNEMINP; nucleotides encoding NEMINP or VPAEGV; or nucleotides 151  
30 to 444 encoding the SH2 domain), or based on all or part of the nucleic acid sequence shown in SEQ ID NO: 7 or Figure 10, and using this labelled nucleic acid probe to screen an appropriate DNA library (e.g. a cDNA or genomic DNA library). For instance, a cDNA library made from hemopoietic cells can be used to isolate a cDNA encoding a protein having SHIP activity by screening the library with the labelled probe using standard techniques. Alternatively, a  
35 genomic DNA library can be similarly screened to isolate a genomic clone encompassing a gene encoding a protein having SH2-containing inositol-phosphatase activity. Nucleic acids isolated by screening of a cDNA or genomic DNA library can be sequenced by standard techniques.

An isolated and purified nucleic acid molecule of the invention which is DNA can also be isolated by selectively amplifying a nucleic acid encoding SHIP using the polymerase chain reaction (PCR) methods and cDNA or genomic DNA. It is possible to design synthetic oligonucleotide primers from the nucleotide sequence shown in SEQ ID NO:1 or Figure 3, or shown in SEQ ID NO:7 or Figure 10, for use in PCR. A nucleic acid can be amplified from cDNA or genomic DNA using these oligonucleotide primers and standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. It will be appreciated that cDNA may be prepared from mRNA, by isolating total cellular mRNA by a variety of techniques, for example, by using the guanidinium-thiocyanate extraction procedure of Chirgwin et al., *Biochemistry*, 18, 5294-5299 (1979). cDNA is then synthesized from the mRNA using reverse transcriptase (for example, Moloney MLV reverse transcriptase available from Gibco/BRL, Bethesda, MD, or AMV reverse transcriptase available from Seikagaku America, Inc., St. Petersburg, FL).

An isolated and purified nucleic acid molecule of the invention which is RNA can be isolated by cloning a cDNA encoding SHIP into an appropriate vector which allows for transcription of the cDNA to produce an RNA molecule which encodes a protein which exhibits phosphoInS-5-ptase activity. For example, a cDNA can be cloned downstream of a bacteriophage promoter, (e.g. a T7 promoter) in a vector, cDNA can be transcribed in vitro with T7 polymerase, and the resultant RNA can be isolated by standard techniques.

A nucleic acid molecule of the invention may also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071).

Determination of whether a particular nucleic acid molecule encodes a protein having SHIP activity can be accomplished by expressing the cDNA in an appropriate host cell by standard techniques, and testing the ability of the expressed protein to associate with Shc and/or hydrolyze a substrate as described herein. A cDNA having the biological activity of SHIP so isolated can be sequenced by standard techniques, such as dideoxynucleotide chain termination or Maxam-Gilbert chemical sequencing, to determine the nucleic acid sequence and the predicted amino acid sequence of the encoded protein.

The initiation codon and untranslated sequences of SHIP or a SHIP related protein may be determined using currently available computer software designed for the purpose, such as PC/Gene (IntelliGenetics Inc., Calif.). The intron-exon structure and the transcription regulatory sequences of the gene encoding the SHIP protein may be identified by using a nucleic acid molecule of the invention encoding SHIP to probe a genomic DNA clone library. Regulatory elements can be identified using conventional techniques. The function of the

elements can be confirmed by using these elements to express a reporter gene such as the bacterial gene lacZ which is operatively linked to the elements. These constructs may be introduced into cultured cells using standard procedures or into non-human transgenic animal models. In addition to identifying regulatory elements in DNA, such constructs may also be used to identify nuclear proteins interacting with the elements, using techniques known in the art.

The 5' untranslated region of murine SHIP comprises nucleotides 1 to 138 in Figure 2(A) or SEQ ID. NO. 1, and the 5' untranslated region of human SHIP comprises nucleotides 1 to 128 in Figure 10 or SEQ ID. NO. 7.

The sequence of a nucleic acid molecule of the invention may be inverted relative to its normal presentation for transcription to produce an antisense nucleic acid molecule. An antisense nucleic acid molecule may be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art.

## II. SHIP Proteins of the Invention

The amino acid sequence of murine SHIP is shown in SEQ.ID.No.2 or in Figure 2 (A) and the amino acid sequence of human SHIP is shown in SEQ.ID.No. 8 or in Figure 11. SHIP contains a number of well-characterized regions including an amino terminal src homology 2 (SH2) domain containing the sequence DGSFLVR which is highly conserved among SH2 domains; two phosphotyrosine binding (PTB) consensus sequences; proline rich regions near the carboxy terminus containing a class I sequence (PPSQPPLSP) and class II sequences (PVKPSR, PPLSPKK, AND PPLPVK); and two motifs highly conserved among inositol polyphosphate-5-phosphatases (i.e. the sequences WLGLNRYR and KYNLPSWCDRLW).

The SHIP protein is expressed in many cell types including hemopoietic cells, bone marrow, lung, spleen, muscles, testes, and kidney.

In addition to the full length SHIP amino acid sequence (SEQ. ID.NO:2 or Figure 2(A); SEQ. ID.NO:8 or Figure 11), the proteins of the present invention include truncations of SHIP, and analogs, and homologs of SHIP and truncations thereof as described herein. Truncated proteins may comprise peptides of between 3 and 1090 amino acid residues, ranging in size from a tripeptide to a 1090 mer polypeptide. For example, a truncated protein may comprise the SH2 domain (the amino acids encoded by nucleotides 151 to 444 as shown in Figure 3 and encoded by nucleotides 141 to 434 in Figure 10); the proline rich regions (the amino acids encoded by nucleotides 3013 to 3580 in Figure 3 and encoded by nucleotides 3009 to 3564 in Figure 10); the 5-phosphatase motifs (amino acids encoded by nucleotides 1886 to 1934 and 2144 to 2167 in Figure 3 and encoded by nucleotides 1876 to 1924 and 2134 to 2157 in Figure 10); the 5-ptase domain (the amino acids encoded by nucleotides 1783 to 2130 in Figure 3 and encoded by nucleotides 1773 to 2120 in Figure 10); the PTB domain target sequences, INPNY and ENPLY (the amino acids encoded by nucleotides 2866-2880 and 3175 to 3189 in Figure 3 and encoded by nucleotides 2856 to 2870 and 3177 to 3191 in Figure 10)); or NPXY sequence of SHIP.

The truncated proteins may have an amino group (-NH<sub>2</sub>), a hydrophobic group (for example, carbobenzoxy, dansyl, or T-butyloxycarbonyl), an acetyl group, a 9-fluorenylmethoxy-carbonyl (PMOC) group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the amino terminal end.

5 The truncated proteins may have a carboxyl group, an amido group, a T-butyloxycarbonyl group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the carboxy terminal end. An isoprenoid may also be attached to a truncated protein comprising the 5-ptase domain to localize SHIP 5-ptase to the inside of the plasma membrane.

10 The proteins of the invention may also include analogs of SHIP as shown in SEQ. ID. NO. 2 or Figure 2 (A), or as shown in SEQ. ID. NO. 8 or Figure 11, and/or truncations thereof as described herein, which may include, but are not limited to, SHIP (SEQ. ID. NO. 2 or Figure 2(A); SEQ. ID. NO. 8 or Figure 11), containing one or more amino acid substitutions, insertions, and/or deletions. Amino acid substitutions may be of a conserved or non-conserved nature.

15 Conserved amino acid substitutions involve replacing one or more amino acids of the SHIP amino acid sequence with amino acids of similar charge, size, and/or hydrophobicity characteristics. When only conserved substitutions are made the resulting analog should be functionally equivalent to SHIP (SEQ. ID. NO. 2 or Figure 2(A); SEQ. ID. NO. 8 or Figure 11). Non-conserved substitutions involve replacing one or more amino acids of the SHIP amino acid

20 sequence with one or more amino acids which possess dissimilar charge, size, and/or hydrophobicity characteristics. By way of example, D675 may be replaced with A675 in Figure 2(A) (or 672 in Figure 11) to create an analog which does not have 5-ptase activity.

One or more amino acid insertions may be introduced into SHIP (SEQ. ID. NO. 2 or Figure 2(A); SEQ. ID. NO. 8 or Figure 11). Amino acid insertions may consist of single amino

25 acid residues or sequential amino acids ranging from 2 to 15 amino acids in length. For example, amino acid insertions may be used to destroy the PTB domain target sequences or the proline-rich consensus sequences so that SHIP can no longer bind SH3-containing proteins.

Deletions may consist of the removal of one or more amino acids, or discrete portions (e.g. one or more of the SH2 domain, PTB consensus sequences; the sequences conserved among

30 inositol polyphosphate-5-phosphatases) from the SHIP (SEQ. ID. NO. 2 or Figure 2(A), SEQ. ID. NO. 8 or Figure 11) sequence. The deleted amino acids may or may not be contiguous. The lower limit length of the resulting analog with a deletion mutation is about 10 amino acids, preferably 100 amino acids.

It is anticipated that if amino acids are replaced, inserted or deleted in sequences

35 outside the amino terminal src homology 2 (SH2) domain, the phosphotyrosine binding (PTB) consensus sequences, the proline rich region and motifs highly conserved among inositol polyphosphate-5-phosphatases, that the resulting analog of SHIP will associate with Shc and exhibit phosphoinos-5-ptase activity.

The proteins of the invention also include homologs of SHIP (SEQ. ID. NO. 2 or Figure 2(A); SEQ. ID. NO. 8 or Figure 11) and/or truncations thereof as described herein. Homology or identity refers to sequence similarity between sequences and it may be determined by comparing a position in each sequence which may be aligned for purposes of comparison. A degree of homology between sequences is a function of the number of matching positions shared by the sequences. Homologs will generally have the same regions which are characteristic of SHIP, namely an amino terminal src homology 2 (SH2) domain, two phosphotyrosine binding (PTB) consensus sequences, a proline rich region and two motifs highly conserved among inositol polyphosphate-5-phosphatases. It is anticipated that, outside of the well-characterized regions of SHIP specified herein (i.e. SH2 domain, PTB domain etc), a protein comprising an amino acid sequence which is about 50% similar, preferably 80 to 90% similar, with the amino acid sequences shown in SEQ ID NO:2 or Figure 2(A), or SEQ. ID. NO. 8 or Figure 11, will exhibit phosphoIns-5-ptase activity and associate with Shc.

A comparison of the amino acid sequences of murine and human SHIP are shown in Figure 12. As shown in Figure 12, human and murine SHIP are 87.2% identical at the amino acid level.

The invention also contemplates isoforms of the protein of the invention. An isoform contains the same number and kinds of amino acids as the protein of the invention, but the isoform has a different molecular structure. The isoforms contemplated by the present invention are those having the same properties as the protein of the invention as described herein.

The present invention also includes SHIP or a SHIP related protein conjugated with a selected protein, or a selectable marker protein (see below) to produce fusion proteins. Further, the present invention also includes activated or phosphorylated SHIP proteins of the invention. Additionally, immunogenic portions of SHIP and SHIP related proteins are within the scope of the invention.

SHIP and SHIP related proteins of the invention may be prepared using recombinant DNA methods. Accordingly, the nucleic acid molecules of the present invention having a sequence which encodes SHIP or a SHIP related protein of the invention may be incorporated in a known manner into an appropriate expression vector which ensures good expression of the protein. Possible expression vectors include but are not limited to cosmids, plasmids, or modified viruses (e.g. replication defective retroviruses, adenoviruses and adeno-associated viruses), so long as the vector is compatible with the host cell used. The expression vectors are "suitable for transformation of a host cell", means that the expression vectors contain a nucleic acid molecule of the invention and regulatory sequences selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid molecule. Operatively linked is intended to mean that the nucleic acid is linked to regulatory sequences in a manner which allows expression of the nucleic acid.

The invention therefore contemplates a recombinant expression vector of the invention containing a nucleic acid molecule of the invention, or a fragment thereof, and the necessary regulatory sequences for the transcription and translation of the inserted protein sequence. Suitable regulatory sequences may be derived from a variety of sources, including bacterial, 5 fungal, viral, mammalian, or insect genes (For example, see the regulatory sequences described in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Selection of appropriate regulatory sequences is dependent on the host cell chosen as discussed below, and may be readily accomplished by one of ordinary skill in the art. Examples of such regulatory sequences include: a transcriptional promoter and enhancer or 10 RNA polymerase binding sequence, a ribosomal binding sequence, including a translation initiation signal. Additionally, depending on the host cell chosen and the vector employed, other sequences, such as an origin of replication, additional DNA restriction sites, enhancers, and sequences conferring inducibility of transcription may be incorporated into the expression vector. It will also be appreciated that the necessary regulatory sequences may be supplied by 15 the native SHIP and/or its flanking regions.

The invention further provides a recombinant expression vector comprising a DNA nucleic acid molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression, by transcription of the DNA molecule, or an RNA 20 molecule which is antisense to the nucleotide sequence of SEQ ID NO: 1 or Figure 2(A), or SEQ. ID. NO. 8 or Figure 11. Regulatory sequences operatively linked to the antisense nucleic acid can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance a viral promoter and/or enhancer, or regulatory sequences can be chosen which direct tissue or cell type specific expression of antisense RNA.

25 The recombinant expression vectors of the invention may also contain a selectable marker gene which facilitates the selection of host cells transformed or transfected with a recombinant molecule of the invention. Examples of selectable marker genes are genes encoding a selectable marker protein such as G418 and hygromycin which confer resistance to certain drugs,  $\beta$ -galactosidase, chloramphenicol acetyltransferase, firefly luciferase, or an 30 immunoglobulin or portion thereof such as the Fc portion of an immunoglobulin preferably IgG. Transcription of the selectable marker gene is monitored by changes in the concentration of the selectable marker protein such as  $\beta$ -galactosidase, chloramphenicol acetyltransferase, or firefly luciferase. If the selectable marker gene encodes a protein conferring antibiotic resistance such as neomycin resistance transformant cells can be selected with G418. Cells that 35 have incorporated the selectable marker gene will survive, while the other cells die. This makes it possible to visualize and assay for expression of recombinant expression vectors of the invention and in particular to determine the effect of a mutation on expression and phenotype.

It will be appreciated that selectable markers can be introduced on a separate vector from the nucleic acid of interest.

The recombinant expression vectors may also contain genes which encode a fusion moiety which provides increased expression of the recombinant protein; increased solubility of the recombinant protein; and aid in the purification of the target recombinant protein by acting as a ligand in affinity purification. For example, a proteolytic cleavage site may be added to the target recombinant protein to allow separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Typical fusion expression vectors include pGEX (Amrad Corp., Melbourne, Australia), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the recombinant protein.

Recombinant expression vectors can be introduced into host cells to produce a transformant host cell. The term "transformant host cell" is intended to include prokaryotic and eukaryotic cells which have been transformed or transfected with a recombinant expression vector of the invention. The terms "transformed with", "transfected with", "transformation" and "transfection" are intended to encompass introduction of nucleic acid (e.g. a vector) into a cell by one of many possible techniques known in the art. Prokaryotic cells can be transformed with nucleic acid by, for example, electroporation or calcium-chloride mediated transformation. Nucleic acid can be introduced into mammalian cells via conventional techniques such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofectin, electroporation or microinjection. Suitable methods for transforming and transfecting host cells can be found in Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory textbooks.

Suitable host cells include a wide variety of prokaryotic and eukaryotic host cells. For example, the proteins of the invention may be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus), yeast cells or mammalian cells. Other suitable host cells can be found in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1991).

More particularly, bacterial host cells suitable for carrying out the present invention include *E. coli*, *B. subtilis*, *Salmonella typhimurium*, and various species within the genus *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, as well as many other bacterial species well known to one of ordinary skill in the art. Suitable bacterial expression vectors preferably comprise a promoter which functions in the host cell, one or more selectable phenotypic markers, and a bacterial origin of replication. Representative promoters include the  $\beta$ -lactamase (penicillinase) and lactose promoter system (see Chang et al., Nature 275:615, 1978), the *trp* promoter (Nichols and Yanofsky, Meth in Enzymology 101:155, 1983) and the *tac* promoter (Russell et al., Gene 20: 231, 1982). Representative selectable markers include various

antibiotic resistance markers such as the kanamycin or ampicillin resistance genes. Suitable expression vectors include but are not limited to bacteriophages such as lambda derivatives or plasmids such as pBR322 (see Bolivar et al., Gene 2:95, 1977), the pUC plasmids pUC18, pUC19, pUC118, pUC119 (see Messing, Meth in Enzymology 101:20-77, 1983 and Vieira and  
5 Messing, Gene 19:259-268, 1982), and pNH8A, pNH16a, pNH18a, and Bluescript M13 (Stratagene, La Jolla, Calif.). Typical fusion expression vectors which may be used are discussed above, e.g. pGEX (Amrad Corp., Melbourne, Australia), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ). Examples of inducible non-fusion expression vectors include pTrc (Amann et al., (1988) Gene 69:301-315) and pET 11d  
10 (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 60-89).

Yeast and fungi host cells suitable for carrying out the present invention include, but are not limited to *Saccharomyces cerevisiae*, the genera *Pichia* or *Kluyveromyces* and various species of the genus *Aspergillus*. Examples of vectors for expression in yeast *S. cerevisiae*  
15 include pYepSec1 (Baldari et al., (1987) Embo J. 6:229-234), pMFa (Kurjan and Herskowitz, (1982) Cell 30:933-943), pJRY88 (Schultz et al., (1987) Gene 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Protocols for the transformation of yeast and fungi are well known to those of ordinary skill in the art. (see Hinnen et al., PNAS USA 75:1929, 1978; Itoh et al., J. Bacteriology 153:163, 1983, and Cullen et al. (Bio/Technology 5:369, 1987).

Mammalian cells suitable for carrying out the present invention include, among others: COS (e.g., ATCC No. CRL 1650 or 1651), BHK (e.g., ATCC No. CRL 6281), CHO (ATCC No. CCL 61), HeLa (e.g., ATCC No. CCL 2), 293 (ATCC No. 1573) and NS-1 cells. Suitable expression vectors for directing expression in mammalian cells generally include a promoter (e.g., derived from viral material such as polyoma, Adenovirus 2, cytomegalovirus and Simian  
25 Virus 40), as well as other transcriptional and translational control sequences. Examples of mammalian expression vectors include pCDM8 (Seed, B., (1987) Nature 329:840) and pMT2PC (Kaufman et al. (1987), EMBOJ. 6:187-195).

Given the teachings provided herein, promoters, terminators, and methods for introducing expression vectors of an appropriate type into plant, avian, and insect cells may  
30 also be readily accomplished. For example, within one embodiment, the proteins of the invention may be expressed from plant cells (see Sinkar et al., J. Biosci (Bangalore) 11:47-58, 1987, which reviews the use of *Agrobacterium rhizogenes* vectors; see also Zambryski et al., Genetic Engineering, Principles and Methods, Hollaender and Setlow (eds.), Vol. VI, pp. 253-278, Plenum Press, New York, 1984, which describes the use of expression vectors for plant  
35 cells, including, among others, pAS2022, pAS2023, and pAS2034).

Insect cells suitable for carrying out the present invention include cells and cell lines from *Bombyx* or *Spodoptera* species. Baculovirus vectors available for expression of proteins in cultured insect cells (SF 9 cells) include the pAc series (Smith et al., (1983) Mol. Cell Biol.

3:2156-2165) and the pVL series (Lucklow, V.A., and Summers, M.D., (1989) Virology 170:31-39).

Alternatively, the proteins of the invention may also be expressed in non-human transgenic animals such as, rats, rabbits, sheep and pigs (see Hammer et al. (Nature 5 315:680-683, 1985), Palmiter et al. (Science 222:809-814, 1983), Brinster et al. (Proc Natl. Acad. Sci USA 82:44384442, 1985), Palmiter and Brinster (Cell. 41:343-345, 1985) and U.S. Patent No. 4,736,866).

The proteins of the invention may also be prepared by chemical synthesis using techniques well known in the chemistry of proteins such as solid phase synthesis (Merrifield, 10 1964, J. Am. Chem. Assoc. 85:2149-2154) or synthesis in homogenous solution (Houbenweyl, 1987, Methods of Organic Chemistry, ed. E. Wansch, Vol. 15 I and II, Thieme, Stuttgart).

N-terminal or C-terminal fusion proteins comprising SHIP or a SHIP related protein of the invention conjugated with other molecules, such as proteins may be prepared by fusing, through recombinant techniques, the N-terminal or C-terminal of SHIP or a SHIP related 15 protein, and the sequence of a selected protein or selectable marker protein with a desired biological function. The resultant fusion proteins contain SHIP or a SHIP related protein fused to the selected protein or marker protein as described herein. Examples of proteins which may be used to prepare fusion proteins include immunoglobulins, glutathione-S-transferase (GST), hemagglutinin (HA), and truncated myc. The present inventor has made GST fusion proteins 20 containing the SH2 domain of SHIP and GST fusion proteins containing the 5-ptase domain attached to an isoprenoid to localize SHIP 5-ptase to the inside of the plasma membrane.

Phosphorylated or activated SHIP or SHIP related proteins of the invention may be prepared using the method described in Reedijk et al. The EMBO Journal 11(4):1365, 1992. For example, tyrosine phosphorylation may be induced by infecting bacteria harbouring a plasmid 25 containing a nucleotide sequence of the invention, with a  $\lambda$ gt11 bacteriophage encoding the cytoplasmic domain of the Elk tyrosine kinase as an Elk fusion protein. Bacteria containing the plasmid and bacteriophage as a lysogen are isolated. Following induction of the lysogen, the expressed protein becomes phosphorylated by the tyrosine kinase.

#### **IV. Utility of the Nucleic Acid Molecules and Proteins of the Invention**

30 The nucleic acid molecules of the invention allow those skilled in the art to construct nucleotide probes for use in the detection of nucleic acid sequences in biological materials. Suitable probes include nucleic acid molecules based on nucleic acid sequences encoding at least 6 sequential amino acids from regions of the SHIP protein as shown in SEQ.ID NO:2 or Figure 2 (A), and SEQ.ID NO:8 or Figure 11. For example, a probe may be based on the nucleotides 2830 2874 in Figure 3 (or SEQ ID.NO. 1) encoding VPAEGVSSLNEMINP; the nucleotides encoding NEMINP or VPAEGV; or the nucleotides 151 to 445 in Figure 3 (or SEQ ID.NO. 1) encoding the SH2 domain. Preferably, the probe comprises a 1 to 1.5kb segment corresponding to the 5' and 3' ends of the 5Kb SHIP mRNA. A nucleotide probe may be labelled with a detectable

substance such as a radioactive label which provides for an adequate signal and has sufficient half-life such as  $^{32}\text{P}$ ,  $^3\text{H}$ ,  $^{14}\text{C}$  or the like. Other detectable substances which may be used include antigens that are recognized by a specific labelled antibody, fluorescent compounds, enzymes, antibodies specific for a labelled antigen, and luminescent compounds. An appropriate label may be selected having regard to the rate of hybridization and binding of the probe to the nucleotide to be detected and the amount of nucleotide available for hybridization. Labelled probes may be hybridized to nucleic acids on solid supports such as nitrocellulose filters or nylon membranes as generally described in Sambrook et al, 1989, Molecular Cloning, A Laboratory Manual (2nd ed.). The nucleic acid probes may be used to detect genes, preferably in human cells, that encode SHIP, and SHIP related proteins. The nucleotide probes may therefore be useful in the diagnosis of disorders of the hemopoietic system including chronic myelogenous leukemia, and acute lymphocytic leukemia, etc.

SHIP or a SHIP related protein of the invention can be used to prepare antibodies specific for the proteins. Antibodies can be prepared which bind a distinct epitope in an unconserved region of the protein. An unconserved region of the protein is one which does not have substantial sequence homology to other proteins, for example the regions outside the well-characterized regions of SHIP as described herein. Alternatively, a region from one of the well-characterized domains (e.g. SH2 domain) can be used to prepare an antibody to a conserved region of SHIP or a SHIP related protein. Antibodies having specificity for SHIP or a SHIP related protein may also be raised from fusion proteins created by expressing for example, trpE-SHIP fusion proteins in bacteria as described herein.

Conventional methods can be used to prepare the antibodies. For example, by using a peptide of SHIP or a SHIP related protein, polyclonal antisera or monoclonal antibodies can be made using standard methods. A mammal, (e.g., a mouse, hamster, or rabbit) can be immunized with an immunogenic form of the peptide which elicits an antibody response in the mammal. Techniques for conferring immunogenicity on a peptide include conjugation to carriers or other techniques well known in the art. For example, the peptide can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassay procedures can be used with the immunogen as antigen to assess the levels of antibodies. Following immunization, antisera can be obtained and, if desired, polyclonal antibodies isolated from the sera.

To produce monoclonal antibodies, antibody producing cells (lymphocytes) can be harvested from an immunized animal and fused with myeloma cells by standard somatic cell fusion procedures thus immortalizing these cells and yielding hybridoma cells. Such techniques are well known in the art, (e.g., the hybridoma technique originally developed by Kohler and Milstein (Nature 256, 495-497 (1975)) as well as other techniques such as the human B-cell hybridoma technique (Kozbor et al., Immunol. Today 4, 72 (1983)), the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al. Monoclonal

Antibodies in Cancer Therapy (1985) Allen R. Bliss, Inc., pages 77-96), and screening of combinatorial antibody libraries (Huse et al., Science 246, 1275 (1989)). Hybridoma cells can be screened immunochemically for production of antibodies specifically reactive with the peptide and the monoclonal antibodies can be isolated. Therefore, the invention also  
5 contemplates hybridoma cells secreting monoclonal antibodies with specificity for SHIP or a SHIP related protein as described herein.

The term "antibody" as used herein is intended to include fragments thereof which also specifically react with a protein, or peptide thereof, having the activity of SHIP. Antibodies can be fragmented using conventional techniques and the fragments screened for  
10 utility in the same manner as described above. For example, F(ab')<sub>2</sub> fragments can be generated by treating antibody with pepsin. The resulting F(ab')<sub>2</sub> fragment can be treated to reduce disulfide bridges to produce Fab' fragments.

Chimeric antibody derivatives, i.e., antibody molecules that combine a non-human animal variable region and a human constant region are also contemplated within the scope of  
15 the invention. Chimeric antibody molecules can include, for example, the antigen binding domain from an antibody of a mouse, rat, or other species, with human constant regions. Conventional methods may be used to make chimeric antibodies containing the immunoglobulin variable region which recognizes the gene product of SHIP antigens of the invention (See, for example, Morrison et al., Proc. Natl Acad. Sci. U.S.A. 81,6851 (1985); Takeda et al., Nature  
20 314, 452 (1985), Cabilly et al., U.S. Patent No. 4,816,567; Boss et al., U.S. Patent No. 4,816,397; Tanaguchi et al., European Patent Publication EP171496; European Patent Publication 0173494, United Kingdom patent GB 2177096B). It is expected that chimeric antibodies would be less immunogenic in a human subject than the corresponding non-chimeric antibody.

25 Monoclonal or chimeric antibodies specifically reactive with a protein of the invention as described herein can be further humanized by producing human constant region chimeras, in which parts of the variable regions, particularly the conserved framework regions of the antigen-binding domain, are of human origin and only the hypervariable regions are of non-human origin. Such immunoglobulin molecules may be made by techniques known in  
30 the art, (e.g., Teng et al., Proc. Natl. Acad. Sci. U.S.A., 80, 7308-7312 (1983); Kozbor et al., Immunology Today, 4, 7279 (1983); Olsson et al., Meth. Enzymol., 92, 3-16 (1982)), and PCT Publication WO92/06193 or EP 0239400). Humanized antibodies can also be commercially produced (Scotgen Limited, 2 Holly Road, Twickenham, Middlesex, Great Britain.)

Specific antibodies, or antibody fragments, reactive against proteins of the invention  
35 may also be generated by screening expression libraries encoding immunoglobulin genes, or portions thereof, expressed in bacteria with peptides produced from the nucleic acid molecules of the present invention. For example, complete Fab fragments, VH regions and FV regions can be expressed in bacteria using phage expression libraries (See for example Ward et al., Nature

341, 544-546: (1989); Huse et al., Science 246, 1275-1281 (1989); and McCafferty et al. Nature 348, 552-554 (1990)). Alternatively, a SCID-hu mouse, for example the model developed by Genpharm, can be used to produce antibodies, or fragments thereof.

Antibodies specifically reactive with SHIP or a SHIP related protein, or derivatives thereof, such as enzyme conjugates or labeled derivatives, may be used to detect SHIP in various biological materials, for example they may be used in any known immunoassays which rely on the binding interaction between an antigenic determinant of SHIP or a SHIP related protein, and the antibodies. Examples of such assays are radioimmunoassays, enzyme immunoassays (e.g.ELISA), immunofluorescence, immunoprecipitation, latex agglutination, hemagglutination, and histochemical tests. Thus, the antibodies may be used to detect and quantify SHIP in a sample in order to determine its role in particular cellular events or pathological states, and to diagnose and treat such pathological states.

In particular, the antibodies of the invention may be used in immuno-histochemical analyses, for example, at the cellular and sub-subcellular level, to detect SHIP, to localise it to particular cells and tissues and to specific subcellular locations, and to quantitate the level of expression.

Cytochemical techniques known in the art for localizing antigens using light and electron microscopy may be used to detect SHIP. Generally, an antibody of the invention may be labelled with a detectable substance and SHIP may be localised in tissue based upon the presence of the detectable substance. Examples of detectable substances include various enzymes, fluorescent materials, luminescent materials and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, biotin, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; and examples of suitable radioactive material include radioactive iodine I125, I131 or tritium. Antibodies may also be coupled to electron dense substances, such as ferritin or colloidal gold, which are readily visualised by electron microscopy.

Indirect methods may also be employed in which the primary antigen-antibody reaction is amplified by the introduction of a second antibody, having specificity for the antibody reactive against SHIP. By way of example, if the antibody having specificity against SHIP is a rabbit IgG antibody, the second antibody may be goat anti-rabbit gamma-globulin labelled with a detectable substance as described herein.

Where a radioactive label is used as a detectable substance, SHIP may be localized by radioautography. The results of radioautography may be quantitated by determining the density of particles in the radioautographs by various optical methods, or by counting the grains.

As discussed herein, SHIP associates with Shc following cytokine stimulation of hemopoietic cells, and it has a role in regulating proliferation, differentiation, activation and metabolism of cells of the hemopoietic system. Therefore, the above described methods for detecting nucleic acid molecules of the invention and SHIP, can be used to monitor proliferation, differentiation, activation and metabolism of cells of the hemopoietic system by detecting and localizing SHIP and nucleic acid molecules encoding SHIP. It would also be apparent to one skilled in the art that the above described methods may be used to study the developmental expression of SHIP and, accordingly, will provide further insight into the role of SHIP in the hemopoietic system.

SHIP has unique and important roles in the regulation of signalling pathways that control gene expression, cell proliferation, differentiation, activation, and metabolism. This finding permits the identification of substances which affect SHIP regulatory systems and which may be used in the treatment of conditions involving perturbation of signalling pathways. The term "SHIP regulatory system" refers to the interaction of SHIP or a SHIP related protein and Shc or a part thereof, to form a SHIP-Shc complex thereby activating a series of regulatory pathways that control gene expression, cell division, cytoskeletal architecture and cell metabolism. Such pathways include the Ras pathway, the pathway that regulates the breakdown of polyphosphoinositides through phospholipase C, and PI-3-kinase activated pathways, such as the emerging rapamycin-sensitive protein kinase B (PKB/Akt) pathway.

A substance which affects SHIP and accordingly a SHIP regulatory system may be assayed using the above described methods for detecting nucleic acid molecules and SHIP and SHIP related proteins, and by comparing the pattern and level of expression of SHIP or SHIP related proteins in the presence and absence of the substance.

Substances which affect SHIP can also be identified based on their ability to bind to SHIP or a SHIP related protein. Therefore, the invention also provides methods for identifying substances which are capable of binding to SHIP or a SHIP related protein. In particular, the methods may be used to identify substances which are capable of binding to, and in some cases activating (i.e., phosphorylating) SHIP or a SHIP related protein of the invention.

Substances which can bind with SHIP or a SHIP related protein of the invention may be identified by reacting SHIP or a SHIP related protein with a substance which potentially binds to SHIP or a SHIP related protein, under conditions which permit the formation of substance-SHIP or -SHIP related protein complexes and assaying for complexes, for free substance, or for non-complexed SHIP or SHIP related protein, or for activation of SHIP or SHIP related protein. Conditions which permit the formation of substance SHIP or SHIP related protein complexes may be selected having regard to factors such as the nature and amounts of the substance and the protein.

The substance-protein complex, free substance or non-complexed proteins may be isolated by conventional isolation techniques; for example, salting out, chromatography, electrophoresis, gel filtration, fractionation, absorption, polyacrylamide gel electrophoresis, agglutination, or combinations thereof. To facilitate the assay of the components, antibody  
5 against SHIP or SHIP related protein or the substance, or labelled SHIP or SHIP related protein, or a labelled substance may be utilized. The antibodies, proteins, or substances may be labelled with a detectable substance as described above.

Substances which bind to and activate SHIP or a SHIP related protein of the invention may be identified by assaying for phosphorylation of the tyrosine residues of the protein, for  
10 example using antiphosphotyrosine antibodies and labelled phosphorus.

SHIP or SHIP related protein, or the substance used in the method of the invention may be insolubilized. For example, SHIP or SHIP related protein or substance may be bound to a suitable carrier. Examples of suitable carriers are agarose, cellulose, dextran, Sephadex, Sepharose, carboxymethyl cellulose polystyrene, filter paper, ion-exchange resin, plastic  
15 film, plastic tube, glass beads, polyamine-methyl vinyl-ether-maleic acid copolymer, amino acid copolymer, ethylene-maleic acid copolymer, nylon, silk, etc. The carrier may be in the shape of, for example, a tube, test plate, beads, disc, sphere etc.

The insolubilized protein or substance may be prepared by reacting the material with a suitable insoluble carrier using known chemical or physical methods, for example, cyanogen  
20 bromide coupling.

The proteins or substance may also be expressed on the surface of a cell using the methods described herein.

The invention also contemplates a method for assaying for an agonist or antagonist of the binding of SHIP or a SHIP related protein with a substance which is capable of binding  
25 with SHIP or a SHIP related protein. The agonist or antagonist may be an endogenous physiological substance or it may be a natural or synthetic substance. Substances which are capable of binding with SHIP or a SHIP related protein may be identified using the methods set forth herein. In a preferred embodiment, the substance is Shc, or a part of Shc, in particular the SH2 domain of Shc, PTB recognition sequences of Shc, or the region containing Y<sup>317</sup> of Shc  
30 (i.e. amino acids 310 to 322) or an activated form thereof. The nucleic acid sequence and the amino acid sequence of Shc are shown in Figures 7 & 8 (SEQ ID. Nos. 3 and 4), respectively. Shc, or a part of Shc, may be prepared using conventional methods, or they may be prepared as fusion proteins (See Lioubin, M.N. Et al., Mol. Cell. Biol. 14(9):5682, 1994, and Kavanaugh, W. M., and L.T. Williams, Science 266:1862, 1994 for methods for making Shc and Shc fusion  
35 proteins). Shc, or part of Shc may be activated i.e. phosphorylated using the methods described for example by Reedijk et al. (The EMBO Journal, 11(4):1365, 1992) for producing a tyrosine phosphorylated protein. The substance may also be an SH3 containing protein such as

Grb2, or a part of Grb2, in particular the SH3 domain of Grb2. The nucleic acid sequence and the amino acid sequence of Grb2 are shown in Figure 9 (SEQ. ID. 5 and NO. 6, respectively).

Therefore, in accordance with a preferred embodiment, a method is provided which comprises providing a known concentration of SHIP or a SHIP related protein, incubating SHIP  
5 or the SHIP related protein with Shc, or a part of Shc, and a suspected agonist or antagonist under conditions which permit the formation of Shc-SHIP or Shc-SHIP related protein complexes, and assaying for Shc-SHIP or Shc-SHIP related protein complexes, for free Shc, for non-complexed SHIP or SHIP related proteins, or for activation of SHIP or SHIP related proteins. Conditions which permit the formation of Shc-SHIP or Shc-SHIP related protein  
10 complexes and methods for assaying for Shc-SHIP or Shc-SHIP related protein complexes, for free Shc, for non-complexed SHIP or SHIP related protein, or for activation of SHIP or SHIP related protein are described herein.

It will be understood that the agonists and antagonists that can be assayed using the methods of the invention may act on one or more of the binding sites on the protein or substance  
15 including agonist binding sites, competitive antagonist binding sites, non-competitive antagonist binding sites or allosteric sites.

The invention also makes it possible to screen for antagonists that inhibit the effects of an agonist of the interaction of SHIP or a SHIP related protein with a substance which is capable of binding to SHIP or a SHIP related protein. Thus, the invention may be used to assay  
20 for a substance that competes for the same binding site of SHIP or a SHIP related protein.

The methods described above may be used to identifying a substance which is capable of binding to an activated SHIP or SHIP related protein, and to assay for an agonist or antagonist of the binding of activated SHIP or SHIP related protein, with a substance which is capable of binding with activated SHIP or activated SHIP related protein. An activated (i.e.  
25 phosphorylated) SHIP or SHIP related protein may be prepared using the methods described for example in Reedijk et al. The EMBO Journal, 11(4):1365, 1992 for producing a tyrosine phosphorylated protein.

It will also be appreciated that intracellular substances which are capable of binding to SHIP or a SHIP related protein may be identified using the methods described herein. For  
30 example, tyrosine phosphorylated proteins (such as the 97 kd and 75 kd proteins) and non-tyrosine phosphorylated proteins which bind to SHIP or a SHIP related protein may be isolated using the method of the invention, cloned, and sequenced.

The invention also contemplates a method for assaying for the affect of a substance on the phosphoIns-5-ptase activity of SHIP or a SHIP related protein having phosphoIns-5-ptase activity comprising reacting a substrate which is capable of being hydrolyzed by SHIP  
35 or SHIP related protein to produce a hydrolysis product, with a substance which is suspected of affecting the phosphoIns-5-ptase activity of SHIP or a SHIP related protein, under conditions which permit the hydrolysis of the substrate, determining the amount of hydrolysis product,

and comparing the amount of hydrolysis product obtained with the amount obtained in the absence of the substance to determine the affect of the substance on the phosphoIns-5-ptase activity of SHIP or SHIP related proteins. Suitable substrates include phosphatidylinositol trisphosphate (PtdIns-3,4,5-P<sub>3</sub>) and inositol tetraphosphate (Ins-1,3,4,5-P<sub>4</sub>). The former  
5 substrate is hydrolyzed to PtdIns-3,4-P<sub>2</sub> which may be identified by incubation with phosphoIns-4-ptase which converts the bis phosphate product to PtdIns-3-P. The latter is hydrolyzed to Ins-1,3,4-P<sub>3</sub> which is identified by treatment with phosphoIns-1-ptase and phosphoIns-4-ptase. Conditions which permit the hydrolysis of the substrate, may be selected having regard to factors such as the nature and amounts of the substance, substrate,  
10 and the amount of SHIP or SHIP related proteins.

The invention further provides a method for assaying for a substance that affects a SHIP regulatory pathway comprising administering to a non-human animal or to a tissue of an animal, a substance suspected of affecting a SHIP regulatory pathway, and quantitating SHIP or nucleic acids encoding SHIP, or examining the pattern and/or level of expression of SHIP, in  
15 the non-human animal or tissue. SHIP may be quantitated and its expression may be examined using the methods described herein.

The substances identified by the methods described herein, may be used for modulating SHIP regulatory pathways and accordingly may be used in the treatment of conditions involving perturbation of SHIP signalling pathways. In particular, the substances  
20 may be particularly useful in the treatment of disorders of the hemopoietic system such as chronic myelogenous leukemia, and acute lymphocytic leukemia.

SHIP is believed to enhance proliferation. Therefore, inhibitors of SHIP (e.g. truncated or point mutants or anti-sense) may be useful in reversing disorders involving excessive proliferation, and stimulators of SHIP may be useful in the treatment of disorders  
25 requiring stimulation of proliferation. Accordingly, the substances identified using the methods of the invention may be used to stimulate or inhibit cell proliferation associated with disorders including various forms of cancer such as leukemias, lymphomas (Hodgkins and non-Hodgkins), sarcomas, melanomas, adenomas, carcinomas of solid tissue, hypoxic tumors, squamous cell carcinomas of the mouth, throat, larynx, and lung, genitourinary cancers such as  
30 cervical and bladder cancer, hematopoietic cancers, head and neck cancers, and nervous system cancers, benign lesions such as papillomas, arthrosclerosis, angiogenesis, and viral infections, in particular HIV infections; and autoimmune diseases including systemic lupus erythematosus, Wegener's granulomatosis, rheumatoid arthritis, sarcoidosis, polyarthritis, pemphigus, pemphigoid, erythema multiforme, Sjogren's syndrome, inflammatory bowel disease, multiple  
35 sclerosis, myasthenia gravis, keratitis, scleritis, Type I diabetes, insulin-dependent diabetes mellitus, Lupus Nephritis, allergic encephalomyelitis. Substances which stimulate cell proliferation identified using the methods of the invention may be useful in the treatment of conditions involving damaged cells including conditions in which degeneration of tissue occurs

such as arthropathy, bone resorption, inflammatory disease, degenerative disorders of the central nervous system; and for promoting wound healing. The SH2 domain of SHIP has been found to be important for tyrosine phosphorylation, binding to Shc, and for translocation to membranes. The SH2 domain has also been shown to be important in the viability of various haemopoietic cells. Therefore, substances which enhance or inhibit SHIP may affect viability of haemopoietic cells, and they may be useful in preventing or treating conditions requiring enhancement or inhibition of viability of haemopoietic cells.

The substances may be formulated into pharmaceutical compositions for administration to subjects in a biologically compatible form suitable for administration *in vivo*. By "biologically compatible form suitable for administration *in vivo*" is meant a form of the substance to be administered in which any toxic effects are outweighed by the therapeutic effects. The substances may be administered to living organisms including humans, and animals. Administration of a therapeutically active amount of the pharmaceutical compositions of the present invention is defined as an amount effective, at dosages and for periods of time necessary to achieve the desired result. For example, a therapeutically active amount of a substance may vary according to factors such as the disease state, age, sex, and weight of the individual, and the ability of antibody to elicit a desired response in the individual. Dosage regima may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be administered daily or the dose may be proportionally reduced as indicated by the exigencies of the therapeutic situation.

The active substance may be administered in a convenient manner such as by injection (subcutaneous, intravenous, etc.), oral administration, inhalation, transdermal application, or rectal administration. Depending on the route of administration, the active substance may be coated in a material to protect the compound from the action of enzymes, acids and other natural conditions which may inactivate the compound.

The compositions described herein can be prepared by per se known methods for the preparation of pharmaceutically acceptable compositions which can be administered to subjects, such that an effective quantity of the active substance is combined in a mixture with a pharmaceutically acceptable vehicle. Suitable vehicles are described, for example, in Remington's Pharmaceutical Sciences (Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., USA 1985). On this basis, the compositions include, albeit not exclusively, solutions of the substances in association with one or more pharmaceutically acceptable vehicles or diluents, and contained in buffered solutions with a suitable pH and iso-osmotic with the physiological fluids.

The reagents suitable for applying the methods of the invention to identify substances that affect a SHIP regulatory system may be packaged into convenient kits providing the necessary materials packaged into suitable containers. The kits may also include suitable supports useful in performing the methods of the invention.

- 25A -

The invention also provides methods for examining the function of the SHIP protein. Cells, tissues, and non-human animals lacking in *SHIP* expression or partially lacking in *SHIP* expression may be developed using recombinant expression vectors of the invention having specific deletion or insertion mutations in the *SHIP* gene. For example, the PTB recognition sequences, SH2 domain, 5-ptase domain, or proline-rich sequences may be deleted. A

5

recombinant expression vector may be used to inactivate or alter the endogenous gene by homologous recombination, and thereby create a *SHIP* deficient cell, tissue or animal.

Null alleles may be generated in cells, such as embryonic stem cells by deletion mutation. A recombinant *SHIP* gene may also be engineered to contain an insertion mutation which inactivates *SHIP*. Such a construct may then be introduced into a cell, such as an embryonic stem cell, by a technique such as transfection, electroporation, injection etc. Cells lacking an intact *SHIP* gene may then be identified, for example by Southern blotting, Northern Blotting or by assaying for expression of *SHIP* using the methods described herein. Such cells may then be fused to embryonic stem cells to generate transgenic non-human animals deficient in *SHIP*. Germline transmission of the mutation may be achieved, for example, by aggregating the embryonic stem cells with early stage embryos, such as 8 cell embryos, *in vitro*; transferring the resulting blastocysts into recipient females and; generating germline transmission of the resulting aggregation chimeras. Such a mutant animal may be used to define specific cell populations, developmental patterns and *in vivo* processes, normally dependent on *SHIP* expression.

The following non-limiting example are illustrative of the present invention:

#### **EXAMPLES**

The following materials and methods were utilized in the investigations outlined in example 1:

#### **PURIFICATION PROTOCOL**

20 litres of B6SUtA<sub>1</sub> cells, grown to confluence in RPMI containing 10% FCS and 5 ng/ml of GM-CSF, were lysed at 2x10<sup>7</sup> cells/ml with PSB containing 0.5% NP40 (Liu et al., Mol. Cell. Biol. 14, 6926 (1994)) and incubated with GSH-beads bearing GST-Grb2-C-SH3. Bound material was eluted by boiling with 1% SDS, 50 mM Tris-Cl, pH 7.5, and diluted to reduce the SDS to < 0.2% for Amicon YM100, Microcon 30 concentration and 3 rounds of Bio-Sep SEC S3000 (Phenomenex) HPLC to remove GST-Grb2-C-SH3 and other low molecular weight material. Following 2D-PAGE (P.H. O'Farrell, J. Biol. Chem. 250, 4007 (1975)), transfer to a PVDF membrane (Liu et al., Mol. Cell. Biol. 14, 6926 (1994)), and Ponceau S staining, the 145-kD spot was excised and sent to the Harvard Microchemistry Facility for trypsin digestion, C<sub>18</sub> HPLC and amino acid sequencing.

#### **CLONING OF cDNA FOR p145**

Degenerate 3' oligonucleotides were synthesized based on the peptide sequence NEMINP, ie 5' GACATCGATGG(G,A)TT(T,G,A)ATCAT(C,T)TC (A,G)TT-3' to carry out PCR amplification 3' and 5' from a plasmid library of randomly primed B6SUtA<sub>1</sub> cDNA employing 5' PCR primers based on plasmid vector sequence flanking the cDNA insertion site. PCR reactions (Expand™ Long Template PCR System, Boehringer Mannheim) were separated on TAE-agarose gels, transferred to Hybond-N+ Blotting membrane (Amersham) and probed for hybridizing bands with a  $\gamma$ -<sup>32</sup>P-dATP end-labelled degenerate oligonucleotide based on the

upstream, but not overlapping, peptide sequence  
 VPAEGV:5'GTAACGGGT(C,T,A,G)CC(C,T,A,G)GC (C,T,A,G)GA(A,G)G(C,T,A,G)GT-3'. A  
 314 bp hybridizing DNA fragment was identified, gel purified, subcloned into Bluescript KS+,  
 sequenced and the projected translation confirmed to match that of the original amino acid  
 5 sequence obtained with the exception of E→C at amino acid #4: VPACGVSSLNEMINP.  
 Specific primers were synthesized based on the DNA sequence to proceed both 3' and 5' of the  
 314 bp original clone to clone 3 overlapping cDNAs totalling 4047 bp in length and encoding a  
 complete coding sequence of 1190 amino acids. DNA sequence was obtained for both strands  
 (Amplicycle, Perkin Elmer), employing both subcloning and oligomer primers. Data base  
 10 comparisons were performed with the MPSearch program, using the Blitz server operated by  
 the European Molecular Biology Laboratory (Heidelberg, Germany).

#### Determining If p145 Is A PhosphoIns-5-ptase

PtdIns[<sup>32</sup>P]-3,4,5-P<sub>3</sub> was prepared using PtdIns-4,5-P<sub>2</sub> and recombinant PtdIns-3-kinase  
 provided by Dr. L. Williams (Chiron Corp) (17). 5-ptase activity was measured by  
 15 evaporating 30,000 cpm of TLC purified PtdIns[<sup>32</sup>P]-3,4,5-P<sub>3</sub> with 150 ug phosphatidylserine  
 under N<sub>2</sub> and resuspending by sonication in assay buffer. Reaction mixtures (25 µl) containing  
 immunoprecipitate or 5-ptase II, 50 mM Tris-Cl, pH 7.5, 10 mM MgCl<sub>2</sub> and substrate were  
 rocked for 30 min at 37°C. Reactions were stopped and the product separated by TLC (L.A.  
 Norris and P.W. Majerus, J. Biol. Chem. 269, 8716 (1994)). Hydrolysis of [3H]Ins-1,3,4,5-P<sub>4</sub> by  
 20 immunoprecipitates was measured as above in 25 µl containing 16 µM [3H]Ins-1,3,4,5-P<sub>4</sub> (6000  
 cpm/nmol) under conditions where the reaction was linear with time (20 min, 37°C) and  
 enzyme amount (C.A. Mitchell et al., J. Biol. Chem. 264, 8873 (1989)). Proof that the InsP<sub>3</sub>  
 product was [3H]Ins-1,3,4-P<sub>3</sub> was obtained by incubation with recombinant inositol-  
 polyphosphate-4- and 1-phosphatase and the bis phosphate products separated on Dowex-  
 25 formate.

#### LEGENDS FOR FIGURES DISCUSSED IN EXAMPLE 1

Figure 1. The Grb2-C-SH3 domain specifically binds the tyrosine phosphorylated, Shc-  
 associated p145. Lysates prepared from B6SUtA<sub>1</sub> cells (2), treated ± IL-3, were either  
 immunoprecipitated with anti-Shc (Transduction Laboratories), followed by protein A  
 30 Sepharose (lanes 1&2) or incubated with GSH bead bound GST-Grb2-N-SH3 (lanes 3&4) or  
 GSH bead bound GST-Grb2-C-SH3 (lanes 5&6). Proteins were eluted by boiling in SDS sample  
 buffer and subjected to Western analysis using 4G10. For lane 7, lysates from IL-3-stimulated  
 B6SUtA<sub>1</sub> cells were incubated with GSH bead bound GST-Grb2-C-SH3, and anti-Shc  
 immunoprecipitates carried out with the unbound material.

35 Figure 2. Amino acid sequence of p145. (A) Deduced amino acid sequence of p145. The hatched  
 box indicates the SH2 domain; the heavily underlined amino acids, the 2 target sequences for  
 binding to PTB domains; the asterisks, the location of the proline rich motifs; and the lightly  
 underlined amino acids, the 2 conserved 5-ptase motifs. Data base comparisons were

performed with the MPSearch program using the Blitz server operated by the European Molecular Biology Laboratory (Heidelberg, Germany). (B) Diagrammatic representation of the various domains within p145.

**Figure 4. Anti-15<sup>mer</sup> antiserum recognizes the Shc-associated p145 and co-precipitates Shc.**

5 (A) Lysates from B6SUA<sub>1</sub> cells, treated  $\pm$  IL-3, were either immunoprecipitated with anti-Shc (lanes 1&2), NRS (lanes 3&4) or anti-15<sup>mer</sup> (lanes 5&6) or precleared with anti-15<sup>mer</sup> and then immunoprecipitated with anti-Shc (lanes 7&8). Western analysis was then performed with 4G10. (B) Lysates from B6SUA<sub>1</sub> cells, stimulated with IL-3, were immunoprecipitated with anti-Shc or anti-15<sup>mer</sup> and the bound proteins eluted at 23°C for 30 min with SDS-sample  
10 buffer containing 1 mM N-ethylmaleimide in lieu of 2-mercaptoethanol. Western blotting was then carried out with 4G10 (upper panel) and the blot reprobed with anti-Shc (lower panel).

**Figure 5. Expression of p145 RNA in murine tissues.** Northern blot analysis of 2  $\mu$ g of polyA RNA from various tissues probed with a random primer-labeled PCR fragment encompassing a 1.5-kb fragment corresponding to the 3' end of the p145 cDNA (lanes 1-6, spleen, lung, liver,  
15 skeletal muscle, kidney and testes, respectively (Clontech); lane 7, separately prepared blot of bone marrow). Similar intensities were observed upon probing with a random primer-labeled PCR fragment encompassing a 1.5-kb fragment corresponding to the 5' end. Exposure time was 30 hrs. In addition to the prominent 5-kb band, a faint band of 4.5-kb was apparent on the autoradiogram.

20 **Figure 6. p145 contains Ins-1,3,4,5-P<sub>4</sub> and PtdIns-3,4,5-P<sub>3</sub> 5-phosphatase activity.** (A) 2x10<sup>7</sup> B6SUA<sub>1</sub> cells were lysed and anti-15<sup>mer</sup>, anti-Shc and NRS immunoprecipitates incubated with [<sup>3</sup>H]Ins-1,3,4,5-P<sub>4</sub> under conditions where product formation was linear with time. Assays were also carried out  $\pm$  recombinant 5-ptase II as controls. (B) 1/10th of anti-15<sup>mer</sup>, NRS and anti-Shc immunoprecipitates (as well as  $\pm$  recombinant 5-ptase II, ie.  
25 PtlI&BL(blank))) were incubated with PtdIns[<sup>32</sup>P]-3,4,5-P<sub>3</sub> under conditions where product formation was linear with time and the reaction mixture chromatographed on TLC (18).

#### EXAMPLE 1

In preliminary studies aimed at purifying p145, immobilized GST fusion proteins containing the C-terminal (but not the N-terminal) SH3 domain of Grb2 were found to bind a  
30 prominent tyrosine phosphorylated protein doublet from B6SUA<sub>1</sub> cell lysates that possessed the same mobility in SDS-gels as p145 (Figure 1, lanes 1-6). Silver stained gels of Grb2-C-SH3 bound material indicated this doublet was prominent in terms of protein level as well, and most abundant in B6SUA<sub>1</sub> cells (compared to MO7E, TF1, Ba/F3, DA-3 and 32D cells, data not shown). To determine if this Grb2-C-SH3 purified doublet was p145, B6SUA<sub>1</sub> cell lysates  
35 were precleared with Grb2-C-SH3 beads and this dramatically depleted p145 in subsequent anti-Shc immuno-precipitates (Figure 1, lane 7). Further proof was obtained by carrying out 2D-PAGE (P.H. O'Farrell, *J. Biol. Chem.* 250, 4007 (1975)) with the two preparations,

followed by Western analysis, using anti-PY antibodies. An identical pattern of multiple spots was obtained in the 145-kD range, with isoelectric points ranging from 7.2 to 7.8.

Based on these findings, a purification protocol was devised as described above and two sequences were obtained from the purified protein; VPAEGVSSLNEMINP, which was used to construct degenerate oligonucleotides, and DGSFLVR, which strongly suggested the presence of an SH2 domain.

The full length cDNA for p145 was then cloned using a PCR based strategy and a B6SUA<sub>1</sub> cDNA library as described above. The deduced 1190 amino acid sequence, possessing a theoretical pI of 7.75 (consistent with the 2D-gel results) revealed several interesting motifs (Figure 2). Close to the amino terminus is the DGSFLVR sequence that is highly conserved among SH2 domains and, taken together with sequences surrounding this motif, suggests that p145 contains an SH2 domain most homologous, at the protein level, to those within Abl, Bruton's tyrosine kinase and Grb2. There are also two motifs, ie., INPNY and ENPLY, that, in their phosphorylated forms, are theoretically capable of binding to PTB domains (P. Blaikie *et al.*, *J. Biol. Chem.* **269**, 32031 (1994); W.M. Kavanaugh *et al.*, *Science* **268**, 1177 (1995); I. Dikic *et al.*, *J. Biol. Chem.* **270**, 15125 (1995); P. Bork and B. Margolis, *Cell* **80**, 693 (1995); Z. Songyang *et al.*, *J. Biol. Chem.* **270**, 14863 (1995); A. Craparo *et al.*, *J. Biol. Chem.* **270**, 15639 (1995); P. van der Geer and T. Pawson, *TIBS* **20**, 277 (1995); A.G. Batzer *et al.*, *Mol. Cell. Biol.* **15**, 4403 (1995); T. Trub *et al.*, *J. Biol. Chem.* **270**, 18205 (1995)). As well, several predicted proline-rich motifs are present near the carboxy terminus, including both class I (eg, PPSQPPLSP) and class II (eg, PVKPSR, PPLSPKK, PPLPVK (K. Alexandropoulos *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3110 (1995); C. Schumacher *et al.*, *J. Biol. Chem.* **270**, 15341 (1995)). Most interestingly, there are 2 motifs that are highly conserved among 5-ptases, ie, WLGDLYNR and, 73 amino acids C-terminal to this, KYNLPWCDRVLW (X. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4853 (1995)).

To identify tyrosine phosphorylated proteins that interact with p145 in vivo and to confirm p145 had been sequenced, lysates from B6SUA<sub>1</sub> cells were immunoprecipitated with rabbit antiserum (ie, anti-15<sup>mer</sup>) generated against the 15<sup>mer</sup> used for cloning (E. Harlow and D. Lane, *Antibodies, A Laboratory Manual*. Cold Spring Harbor Laboratory, (1988)). Western analysis, using anti-PY, revealed, as expected, a 145-kD tyrosine phosphorylated doublet with an identical mobility in SDS gels to p145 (Figure 4(A), lanes 1&2 and 5&6). Pre-immune serum did not immunoprecipitate this or any other tyrosine phosphorylated protein (Figure 4(A), lanes 3&4). Moreover, anti-Shc immunoprecipitates of lysates precleared with anti-15<sup>mer</sup> no longer contained p145 (Figure 4(A), lane 8). Interestingly, anti-15<sup>mer</sup> immunoprecipitates from lysates of IL-3-stimulated B6SUA<sub>1</sub> cells consistently contained 50-55-kD and, occasionally, 75- and 97-kD tyrosine phosphorylated proteins (Figure 4(A), lane 6). The 50-55-kD protein was shown to be Shc by treating anti-15<sup>mer</sup> immunoprecipitates with N-ethylmaleimide prior to SDS-PAGE to alter the mobility of the interfering IgH chain (M.R.

Block *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 85, 7852 (1988)), and then carrying out Western analysis with anti-PY (Figure 4(B), upper panel) and anti-Shc antibodies (Figure 4(B), lower panel).

To examine whether the expression of p145 was restricted to hemopoietic cells, Northern blot analysis was carried out with polyA purified RNA from various murine tissues. A 5.0-kb p145 transcript was found to be expressed in bone marrow, lung, spleen, muscle, testes and kidney, suggesting the presence of this protein in many cell types (Figure 5).

Lastly, to determine if p145 was indeed a 5-ptase, lysates from B6SUA<sub>1</sub> cells were immunoprecipitated with anti-15<sup>mer</sup>, anti-Shc or normal rabbit serum (NRS) and the immunoprecipitates tested with various 5-ptase substrates (X. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 92:4853 (1995) and as described herein). As can be seen in Figure 6(A), anti-15<sup>mer</sup>, but not NRS, immunoprecipitates hydrolyzed [<sup>3</sup>H]Ins-1,3,4,5-P<sub>4</sub> to [<sup>3</sup>H]Ins-1,3,4-P<sub>3</sub>. The product of the reaction was shown to be [<sup>3</sup>H]Ins-1,3,4-P<sub>3</sub> by incubation with recombinant inositol-polyphosphate-1- and 4-phosphatases, followed by the separation of the bisphosphate product on Dowex-formate (Zhang, X., *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 92:4853-4856, 1995 and Jefferson, A.B. And Majerus, P.W. *J. Biol. Chem.* 270:9370-9377, 1995). In the presence of 3 mM EDTA, no hydrolysis of [<sup>3</sup>H]Ins-1,3,4,5-P<sub>4</sub> was observed, suggesting that this 5-ptase is Mg<sup>++</sup>-dependent. Interestingly, no significant difference in activity was observed between anti-15<sup>mer</sup> immunoprecipitates from stimulated and unstimulated cells. Moreover, as one might expect, anti-Shc immunoprecipitates possessed 5-ptase activity, but only after IL-3-stimulation. In addition, anti-15<sup>mer</sup>, but not NRS, immunoprecipitates catalyzed the hydrolysis of PtdIns[<sup>32</sup>P]-3,4,5-P<sub>3</sub>, as did recombinant 5-ptase II (Figure 6(B)). Once again there was no significant difference in activity between IL-3-stimulated and unstimulated cells and anti-Shc immunoprecipitates possessed 5-ptase activity only after cells were stimulated. This suggests that IL-3 affects only the localization of p145 and not its 5-ptase activity. In studies with other 5-ptase substrates, anti-15<sup>mer</sup> immunoprecipitates did not hydrolyse Ins-1,4,5-P<sub>3</sub> or PtdIns-4,5-P<sub>2</sub>. P145 5-ptase substrate specificity is therefore distinct from that of other 5-ptases such as 5-ptase II, OCRL 5-ptase and a novel Mg<sup>++</sup>-independent 5-ptase (Zhang, X., *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 92:4853-4856, 1995; Jefferson, A.B. And Majerus, P.W. *J. Biol. Chem.* 270:9370-9377, 1995; and Jackson, S.P. *Et al.*, *EMBO J.* 14:4490-4500, 1995).

Of the 5-ptases cloned to date (X. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 92:4853 (1995)), p145 is the first to possess an SH2 domain and to be tyrosine phosphorylated. Thus, p145 may play an important role in cytokine mediated signalling. In this regard, Cullen *et al* just reported that Ins-1,3,4,5-P<sub>4</sub>, which is rapidly elevated in stimulated cells (I.R. Batty *et al.*, *Biochem. J.* 232, 211 (1985)), binds to and stimulates a member of the GAP1 family (P.J. Cullen *et al.*, *Nature* 376, 527 (1995)). It is therefore conceivable that p145, through its association with Shc, regulates Ras activity by hydrolyzing RasGAP bound Ins-1,3,4,5-P<sub>4</sub>. In

addition, with its multiple protein:protein interaction domains and its unique 5-ptase substrate specificity, p145 could play an important role in regulating  $\text{Ca}^{++}$ -independent PKC activity (Toker, A., et al., J. Biol. Chem. 269:32358-32367, 1994), the emerging Akt/PKB pathway (Burgering, B.M. And Coffey, P.J., Nature 376:599-602, 1995 )and other as yet uncharacterized PI-3-kinase stimulated cascades. In terms of its association with Shc, p145 may interact via its phosphorylated tyrosines with the SH2 of Shc, via its phosphorylated PTB recognition sequences with the PTB of Shc (as suggested by *in vitro* studies with the Shc-associated p145 in 3T3 cells ( F.A. Norris and P.W. Majerus, J. Biol. Chem. 269, 8716 (1994)) and/or via its SH2 domain with Y<sup>317</sup> of Shc.

In summary, a tyrosine phosphorylated 145 kDa protein has been purified that associates with Shc in response to multiple cytokines from hemopoietic cells and shown it to be a novel, SH2-containing 5-ptase. Based on its properties it is suggested it be called SHIP for SH2-containing inositol-phosphatase.

## EXAMPLE 2

### Cloning of hSHIP cDNA

Duplicate nitrocellulose (Schleicher & Schuell, Keene, NH) plaque-lifts were prepared from approximately  $1 \times 10^6$  pfu of a custom-made MO7e/MO7-ER  $\lambda$ gt11 cDNA library created from 10 $\mu$ g of poly-A RNA (Clontech, Palo Alto, CA). Phage DNA bound to these membranes was denatured and hybridized (1.5X SSPE, 1% SDS, 1% Blotto, 0.25mg/ml ssDNA) at 50°C for 18 hours with non-overlapping, [ $\lambda^{32}\text{P}$ ]dCTP randomly labeled cDNA fragments corresponding to either 1.5 kb of the 5' - most region (including the SH2 domain) or 1.1 kb of the central region (including the 5-Ptase domain) of murine SHIP. Probed membranes were washed three times with 0.5X SSC, 0.5% SDS at 50°C for 30 minutes each. Membranes were exposed to Kodak X-Omat film (Rochester, NY) and plaques which hybridized with both probes were identified and the phage isolated. Thirteen cDNA inserts were removed from "positive" phage by EcoRI digestion, gel purified, and subcloned into pBluescript KS+ for further analysis. One full-length cDNA, 4926 nt in length, was further digested with either PstI or XhoI and re-subcloned into pBluescript KS+ for automated ABI/Taq Polymerase sequencing (NAPS Unit, University of British Columbia, Vancouver, Canada) using standard T7 and T3 oligoprimers. Regions not overlapped by restriction fragments were sequenced using specific nucleotide oligoprimers. The human SHIP CDNA sequence is set out in Figure 10 and in SEQ.ID.NO.12.

Having illustrated and described the principles of the invention in a preferred embodiment, it should be appreciated to those skilled in the art that the invention can be modified in arrangement and detail without departure from such principles. We claim all modifications coming within the scope of the following claims.

All publications, patents and patent applications referred to herein are incorporated by reference in their entirety to the same extent as if each individual publication, patent or

patent application was specifically and individually indicated to be incorporated by reference in its entirety.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT:
  - (A) NAME: Krystal, Gerald
  - (B) STREET: 601 West 10th Street
  - (C) CITY: Vancouver
  - (D) STATE: British Columbia
  - (E) COUNTRY: Canada
  - (F) POSTAL CODE: V52 1L3
- (ii) TITLE OF INVENTION: SH2-CONTAINING INOSITOL-PHOSPHATASE
- (iii) NUMBER OF SEQUENCES: 8
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: BERESKIN & PARR
  - (B) STREET: 40 KING STREET WEST
  - (C) CITY: TORONTO
  - (D) STATE: ONTARIO
  - (E) COUNTRY: CANADA
  - (F) ZIP: M5H 3Y2
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: PCT/CA96/00655
  - (B) FILING DATE: 27 SEPT 1996
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Kurdydyk, Linda M.
  - (B) REGISTRATION NUMBER: 34,971
  - (C) REFERENCE/DOCKET NUMBER: 7771-018
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 416-364-7311
  - (B) TELEFAX: 416-361-1398

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4040 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: murine
- (vii) IMMEDIATE SOURCE:
  - (B) CLONE: mSHIP
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 139..3693

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCCTGGTAGG AGCAGCAGAG GCAATTTCTG AGAGGCAACA GCGGCAGGT CTCAGCCTAG	60
AGAGGGCCCT GAACTACTTT GCTGGAGTGT CCGTCCTGGG AGTGGCTGCT GACCCAGTCC	120
AGGAGACCCA TGCCTGCC ATG GTC CCT GGG TGG AAC CAT GGC AAC ATC ACC	171
Met Val Pro Gly Trp Asn His Gly Asn Ile Thr	
1 5 10	
CGC TCC AAG GCA GAG GAG CTA CTT TCC AGA GCC GGC AAG GAC GGG AGC	219
Arg Ser Lys Ala Glu Glu Leu Leu Ser Arg Ala Gly Lys Asp Gly Ser	
15 20 25	
TTC CTT GTG CGT GCC AGC GAG TCC ATC CCC CGG GCC TGC GCA CTC TGC	267
Phe Leu Val Arg Ala Ser Glu Ser Ile Pro Arg Ala Cys Ala Leu Cys	
30 35 40	
GTG CTG TTC CGG AAT TGT GTT TAC ACT TAC AGG ATT CTG CCC AAT GAG	315
Val Leu Phe Arg Asn Cys Val Tyr Thr Tyr Arg Ile Leu Pro Asn Glu	
45 50 55	
GAC GAT AAA TTC ACT GTT CAG GCA TCC GAA GGT GTC CCC ATG AGG TTC	363
Asp Asp Lys Phe Thr Val Gln Ala Ser Glu Gly Val Pro Met Arg Phe	
60 65 70 75	
TTC ACG AAG CTG GAC CAG CTC ATC GAC TTT TAC AAG AAG GAA AAC ATG	411
Phe Thr Lys Leu Asp Gln Leu Ile Asp Phe Tyr Lys Lys Glu Asn Met	
80 85 90	
GGG CTG GTG ACC CAC CTG CAG TAC CCC GTG CCC CTG GAG GAG GAG GAT	459
Gly Leu Val Thr His Leu Gln Tyr Pro Val Pro Leu Glu Glu Glu Asp	
95 100 105	
GCT ATT GAT GAG GCT GAG GAG GAC ACT GAA AGT GTC ATG TCA CCA CCT	507
Ala Ile Asp Glu Ala Glu Glu Asp Thr Glu Ser Val Met Ser Pro Pro	
110 115 120	
GAG CTG CCT CCC AGA AAC ATT CCT ATG TCT GCC GGG CCC AGC GAG GCC	555
Glu Leu Pro Pro Arg Asn Ile Pro Met Ser Ala Gly Pro Ser Glu Ala	
125 130 135	
AAG GAC CTT CCT CTT GCA ACA GAG AAC CCC CGA GCC CCT GAG GTC ACC	603
Lys Asp Leu Pro Leu Ala Thr Glu Asn Pro Arg Ala Pro Glu Val Thr	
140 145 150 155	
CGG CTG AGT CTC TCC GAG ACA CTG TTT CAG CGT CTA CAG AGC ATG GAT	651
Arg Leu Ser Leu Ser Glu Thr Leu Phe Gln Arg Leu Gln Ser Met Asp	
160 165 170	
ACC AGT GGG CTT CCC GAG GAG CAC CTG AAA GCC ATC CAG GAT TAT CTG	699
Thr Ser Gly Leu Pro Glu Glu His Leu Lys Ala Ile Gln Asp Tyr Leu	
175 180 185	
AGC ACT CAG CTC CTC CTG GAT TCC GAC TTT TTG AAA ACG GGC TCC AGC	747
Ser Thr Gln Leu Leu Leu Asp Ser Asp Phe Leu Lys Thr Gly Ser Ser	
190 195 200	
AAC CTC CCT CAC CTG AAG AAG CTG ATG TCA CTG CTC TGC AAG GAG CTC	795
Asn Leu Pro His Leu Lys Lys Leu Met Ser Leu Leu Cys Lys Glu Leu	
205 210 215	
CAT GGG GAA GTC ATC AGG ACT CTG CCA TCC CTG GAG TCT CTG CAG AGG	843
His Gly Glu Val Ile Arg Thr Leu Pro Ser Leu Glu Ser Leu Gln Arg	
220 225 230 235	

TTG	TTT	GAC	CAA	CAG	CTC	TCC	CCA	GGC	CTT	CGC	CCA	CGA	CCT	CAG	GTG	891
Leu	Phe	Asp	Gln	Gln	Leu	Ser	Pro	Gly	Leu	Arg	Pro	Arg	Pro	Gln	Val	
			240						245					250		
CCC	GGA	GAG	GCC	AGT	CCC	ATC	ACC	ATG	GTT	GCC	AAA	CTC	AGC	CAA	TTG	939
Pro	Gly	Glu	Ala	Ser	Pro	Ile	Thr	Met	Val	Ala	Lys	Leu	Ser	Gln	Leu	
			255					260					265			
ACA	AGT	CTG	CTG	TCT	TCC	ATT	GAA	GAT	AAG	GTC	AAG	TCC	TTG	CTG	CAC	987
Thr	Ser	Leu	Leu	Ser	Ser	Ile	Glu	Asp	Lys	Val	Lys		280	Leu	Leu	His
		270					275									
GAG	GGC	TCA	GAA	TCT	ACC	AAC	AGG	CGT	TCC	CTT	ATC	CCT	CCG	GTC	ACC	1035
Glu	Gly	Ser	Glu	Ser	Thr	Asn	Arg	Arg	Ser	Leu	Ile	Pro	Pro	Val	Thr	
	285					290					295					
TTT	GAG	GTG	AAG	TCA	GAG	TCC	CTG	GGC	ATT	CCT	CAG	AAA	ATG	CAT	CTC	1083
Phe	Glu	Val	Lys	Ser	Glu	Ser	Leu	Gly	Ile	Pro	Gln	Lys	Met	His	Leu	
300					305					310					315	
AAA	GTG	GAC	GTT	GAG	TCT	GGG	AAA	CTG	ATC	GTT	AAG	AAG	TCC	AAG	GAT	1131
Lys	Val	Asp	Val	Glu	Ser	Gly	Lys	Leu	Ile	Val	Lys	Lys	Ser	Lys	Asp	
				320					325					330		
GGT	TCT	GAG	GAC	AAG	TTC	TAC	AGC	CAC	AAA	AAA	ATC	CTG	CAG	CTC	ATT	1179
Gly	Ser	Glu	Asp	Lys	Phe	Tyr	Ser	His	Lys	Lys	Ile	Leu	Gln	Leu	Ile	
			335					340					345			
AAG	TCC	CAG	AAG	TTT	CTA	AAC	AAG	TTG	GTG	ATT	TTG	GTG	GAG	ACG	GAG	1227
Lys	Ser	Gln	Lys	Phe	Leu	Asn	Lys	Leu	Val	Ile	Leu	Val	Glu	Thr	Glu	
		350					355					360				
AAG	GAG	AAA	ATC	CTG	AGG	AAG	GAA	TAT	GTT	TTT	GCT	GAC	TCT	AAG	AAA	1275
Lys	Glu	Lys	Ile	Leu	Arg	Lys	Glu	Tyr	Val	Phe	Ala	Asp	Ser	Lys	Lys	
	365					370				375						
AGA	GAA	GGC	TTC	TGT	CAA	CTC	CTG	CAG	CAG	ATG	AAG	AAC	AAG	CAT	TCG	1323
Arg	Glu	Gly	Phe	Cys	Gln	Leu	Leu	Gln	Gln	Met	Lys	Asn	Lys	His	Ser	
380				385						390					395	
GAG	CAG	CCA	GAG	CCT	GAC	ATG	ATC	ACC	ATC	TTC	ATT	GGC	ACT	TGG	AAC	1371
Glu	Gln	Pro	Glu	Pro	Asp	Met	Ile	Thr	Ile	Phe	Ile	Gly	Thr	Trp	Asn	
				400					405					410		
ATG	GGT	AAT	GCA	CCC	CCT	CCC	AAG	AAG	ATC	ACG	TCC	TGG	TTT	CTC	TCC	1419
Met	Gly	Asn	Ala	Pro	Pro	Pro	Lys	Lys	Ile	Thr	Ser	Trp	Phe	Leu	Ser	
			415					420					425			
AAG	GGG	CAG	GGA	AAG	ACA	CGG	GAC	GAC	TCT	GCT	GAC	TAC	ATC	CCC	CAT	1467
Lys	Gly	Gln	Gly	Lys	Thr	Arg	Asp	Asp	Ser	Ala	Asp	Tyr	Ile	Pro	His	
		430					435					440				
GAC	ATC	TAT	GTG	ATT	GGC	ACC	CAG	GAG	GAT	CCC	CTT	GGA	GAG	AAG	GAG	1515
Asp	Ile	Tyr	Val	Ile	Gly	Thr	Gln	Glu	Asp	Pro	Leu	Gly	Glu	Lys	Glu	
	445				450						455					
TGG	CTG	GAG	CTA	CTC	AGG	CAC	TCC	CTG	CAA	GAA	GTC	ACC	AGC	ATG	ACA	1563
Trp	Leu	Glu	Leu	Leu	Arg	His	Ser	Leu	Gln	Glu	Val	Thr	Ser	Met	Thr	
460				465						470					475	
TTT	AAA	ACA	GTT	GCC	ATC	CAC	ACC	CTC	TGG	AAC	ATT	CGC	ATA	GTG	GTG	1611
Phe	Lys	Thr	Val	Ala	Ile	His	Thr	Leu	Trp	Asn	Ile	Arg	Ile	Val	Val	
				480					485					490		
CTT	GCC	AAG	CCA	GAG	CAT	GAG	AAT	CGG	ATC	AGC	CAT	ATC	TGC	ACT	GAC	1659
Leu	Ala	Lys	Pro	Glu	His	Glu	Asn	Arg	Ile	Ser	His	Ile	Cys	Thr	Asp	

495				500				505								
AAC Asn	GTG Val	AAG Lys 510	ACA Thr	GGC Gly	ATC Ile	GCC Ala	AAC Asn 515	ACC Thr	CTG Leu	GGA Gly	AAC Asn	AAG Lys 520	GGA Gly	GCA Ala	GTG Val	1707
GGA Gly	GTG Val 525	TCC Ser	TTC Phe	ATG Met	TTC Phe	AAT Asn 530	GGA Gly	ACC Thr	TCC Ser	TTG Leu	GGG Gly 535	TTC Phe	GTC Val	AAC Asn	AGC Ser	1755
CAC His 540	TTG Leu	ACT Thr	TCT Ser	GGA Gly	AGT Ser 545	GAA Glu	AAA Lys	AAG Lys	CTC Leu	AGG Arg 550	AGA Arg	AAT Asn	CAA Gln	AAC Asn	TAT Tyr 555	1803
ATG Met	AAC Asn	ATC Ile	CTG Leu	CGG Arg 560	TTC Phe	CTG Leu	GCC Ala	CTG Leu	GGA Gly 565	GAC Asp	AAG Lys	AAG Lys	CTA Leu	AGC Ser 570	CCA Pro	1851
TTT Phe	AAC Asn	ATC Ile	ACC Thr 575	CAC His	CGC Arg	TTC Phe	ACC Thr	CAC His 580	CTC Leu	TTC Phe	TGG Trp	CTT Leu	GGG Gly 585	GAT Asp	CTC Leu	1899
AAC Asn	TAC Tyr	CGC Arg 590	GTG Val	GAG Glu	CTG Leu	CCC Pro	ACT Thr 595	TGG Trp	GAG Glu	GCA Ala	GAG Glu	GCC Ala 600	ATC Ile	ATC Ile	CAG Gln	1947
AAG Lys	ATC Ile 605	AAG Lys	CAA Gln	CAG Gln	CAG Gln	TAT Tyr 610	TCA Ser	GAC Asp	CTT Leu	CTG Leu	GCC Ala 615	CAC His	GAC Asp	CAA Gln	CTG Leu	1995
CTC Leu 620	CTG Leu	GAG Glu	AGG Arg	AAG Lys	GAC Asp 625	CAG Gln	AAG Lys	GTC Val	TTC Phe	CTG Leu 630	CAC His	TTT Phe	GAG Glu	GAG Glu	GAA Glu 635	2043
GAG Glu	ATC Ile	ACC Thr	TTC Phe	GCC Ala 640	CCC Pro	ACC Thr	TAT Tyr	CGA Arg	TTT Phe 645	GAA Glu	AGA Arg	CTG Leu	ACC Thr	CGG Arg 650	GAC Asp	2091
AAG Lys	TAT Tyr	GCA Ala	TAC Tyr 655	ACG Thr	AAG Lys	CAG Gln	AAA Lys	GCA Ala 660	ACA Thr	GGG Gly	ATG Met	AAG Lys	TAC Tyr 665	AAC Asn	TTG Leu	2139
CCG Pro	TCC Ser	TGG Trp 670	TGC Cys	GAC Asp	CGA Arg	GTC Val	CTC Leu 675	TGG Trp	AAG Lys	TCT Ser	TAC Tyr	CCG Pro 680	CTG Leu	GTG Val	CAT His	2187
GTG Val	GTC Val 685	TGT Cys	CAG Gln	TCC Ser	TAT Tyr	GGC Gly 690	AGT Ser	ACC Thr	AGT Ser	GAC Asp	ATC Ile 695	ATG Met	ACG Thr	AGT Ser	GAC Asp	2235
CAC His 700	AGC Ser	CCT Pro	GTC Val	TTT Phe	GCC Ala 705	ACG Thr	TTT Phe	GAA Glu	GCA Ala	GGA Gly 710	GTC Val	ACA Thr	TCT Ser	CAA Gln	TTC Phe 715	2283
GTC Val	TCC Ser	AAG Lys	AAT Asn	GGT Gly 720	CCT Pro	GGC Gly	ACT Thr	GTA Val	GAT Asp 725	AGC Ser	CAA Gln	GGG Gly	CAG Gln	ATC Ile 730	GAG Glu	2331
TTT Phe	CTT Leu	GCA Ala	TGC Cys 735	TAC Tyr	GCC Ala	ACA Thr	CTG Leu	AAG Lys 740	ACC Thr	AAG Lys	TCC Ser	CAG Gln	ACT Thr 745	AAG Lys	TTC Phe	2379
TAC Tyr	TTG Leu	GAG Glu 750	TTC Phe	CAC His	TCA Ser	AGC Ser	TGC Cys 755	TTA Leu	GAG Glu	AGT Ser	TTT Phe	GTC Val 760	AAG Lys	AGT Ser	CAG Gln	2427

GAA Glu	GGA Gly	GAG Glu	AAT Asn	GAA Glu	GAG Glu	GGA Gly	AGT Ser	GAA Glu	GGA Gly	GAG Glu	CTG Leu	GTG Val	GTA Val	CGG Arg	TTT Phe	2475
765						770					775					
GGA Gly	GAG Glu	ACT Thr	CTT Leu	CCC Pro	AAG Lys	CTA Leu	AAG Lys	CCC Pro	ATT Ile	ATC Ile	TCT Ser	GAC Asp	CCC Pro	GAG Glu	TAC Tyr	2523
780					785					790					795	
TTA Leu	CTG Leu	GAC Asp	CAG Gln	CAT His	ATC Ile	CTG Leu	ATC Ile	AGC Ser	ATT Ile	AAA Lys	TCC Ser	TCT Ser	GAC Asp	AGT Ser	GAC Asp	2571
				800					805					810		
GAG Glu	TCC Ser	TAT Tyr	GGT Gly	GAA Glu	GGC Gly	TGC Cys	ATT Ile	GCC Ala	CTT Leu	CGC Arg	TTG Leu	GAG Glu	ACC Thr	ACA Thr	GAG Glu	2619
			815					820					825			
GCT Ala	CAG Gln	CAT His	CCT Pro	ATC Ile	TAC Tyr	ACG Thr	CCT Pro	CTC Leu	ACC Thr	CAC His	CAT His	GGG Gly	GAG Glu	ATG Met	ACT Thr	2667
		830					835					840				
GGC Gly	CAC His	TTC Phe	AGG Arg	GGA Gly	GAG Glu	ATT Ile	AAG Lys	CTG Leu	CAG Gln	ACC Thr	TCC Ser	CAG Gln	GGC Gly	AAG Lys	ATG Met	2715
	845					850					855					
AGG Arg	GAG Glu	AAG Lys	CTC Leu	TAT Tyr	GAC Asp	TTT Phe	GTG Val	AAG Lys	ACA Thr	GAG Glu	CGG Arg	GAT Asp	GAA Glu	TCC Ser	AGT Ser	2763
860					865					870					875	
GGA Gly	ATG Met	AAA Lys	TGC Cys	TTG Leu	AAG Lys	AAC Asn	CTC Leu	ACC Thr	AGC Ser	CAT His	GAC Asp	CCT Pro	ATG Met	AGG Arg	CAA Gln	2811
				880					885					890		
TGG Trp	GAG Glu	CCT Pro	TCT Ser	GGC Gly	AGG Arg	GTC Val	CCT Pro	GCA Ala	TGT Cys	GGT Gly	GTC Val	TCC Ser	AGC Ser	CTC Leu	AAT Asn	2859
			895					900					905			
GAG Glu	ATG Met	ATC Ile	AAT Asn	CCA Pro	AAC Asn	TAC Tyr	ATT Ile	GGT Gly	ATG Met	GGG Gly	CCT Pro	TTT Phe	GGA Gly	CAG Gln	CCC Pro	2907
		910					915					920				
CTG Leu	CAT His	GGG Gly	AAA Lys	TCA Ser	ACC Thr	CTG Leu	TCC Ser	CCA Pro	GAT Asp	CAG Gln	CAA Gln	CTC Leu	ACA Thr	GCT Ala	TGG Trp	2955
	925					930					935					
AGT Ser	TAT Tyr	GAC Asp	CAG Gln	CTA Leu	CCC Pro	AAA Lys	GAC Asp	TCC Ser	TCC Ser	CTG Leu	GGG Gly	CCT Pro	GGG Gly	AGG Arg	GGG Gly	3003
940					945					950					955	
GAG Glu	GGT Gly	CCT Pro	CCA Pro	ACC Thr	CCT Pro	CCC Pro	TCC Ser	CAA Gln	CCA Pro	CCT Pro	CTG Leu	TCG Ser	CCA Pro	AAG Lys	AAG Lys	3051
				960					965					970		
TTT Phe	TCA Ser	TCT Ser	TCC Ser	ACA Thr	ACC Thr	AAC Asn	CGA Arg	GGT Gly	CCC Pro	TGC Cys	CCC Pro	AGG Arg	GTG Val	CAA Gln	GAG Glu	3099
			975					980					985			
GCA Ala	AGA Arg	CCT Pro	GGG Gly	GAT Asp	CTG Leu	GGA Gly	AAG Lys	GTG Val	GAA Glu	GCT Ala	CTG Leu	CTC Leu	CAG Gln	GAG Glu	GAC Asp	3147
		990				995					1000					
CTG Leu	CTG Leu	CTG Leu	ACG Thr	AAG Lys	CCC Pro	GAG Glu	ATG Met	TTT Phe	GAG Glu	AAC Asn	CCA Pro	CTG Leu	TAT Tyr	GGA Gly	TCC Ser	3195
	1005					1010					1015					
GTG Val	AGT Ser	TCC Ser	TTC Phe	CCT Pro	AAG Lys	CTG Leu	GTG Val	CCC Pro	AGG Arg	AAA Lys	GAG Glu	CAG Gln	GAG Glu	TCT Ser	CCC Pro	3243

1020	1025	1030	1035	
AAG ATG CTG CGG	AAG GAG CCC CCG CCC	TGT CCA GAC CCA GGA	ATC TCA	3291
Lys Met Leu Arg	Lys Glu Pro Pro Pro	Cys Pro Asp Pro Gly	Ile Ser	
	1040	1045	1050	
TCA CCC AGC ATC GTG CTC CCC	AAA GCC CAA GAG GTG GAG	AGT GTC AAG		3339
Ser Pro Ser Ile Val Leu Pro Lys	Ala Gln Glu Val Glu Ser Val	Lys		
	1055	1060	1065	
GGG ACA AGC AAA CAG GCC CCT GTG CCT GTC CTT GGC CCC ACA CCC CGG				3387
Gly Thr Ser Lys Gln Ala Pro Val Pro Val Leu Gly Pro Thr Pro Arg				
	1070	1075	1080	
ATC CGC TCC TTT ACC TGT TCT TCT TCT GCT GAG GGC AGA ATG ACC AGT				3435
Ile Arg Ser Phe Thr Cys Ser Ser Ser Ala Glu Gly Arg Met Thr Ser				
	1085	1090	1095	
GGG GAC AAG AGC CAA GGG AAG CCC AAG GCC TCA GCC AGT TCC CAA GCC				3483
Gly Asp Lys Ser Gln Gly Lys Pro Lys Ala Ser Ala Ser Ser Gln Ala				
	1100	1105	1110	1115
CCA GTG CCA GTC AAG AGG CCT GTC AAG CCT TCC AGG TCA GAA ATG AGC				3531
Pro Val Pro Val Lys Arg Pro Val Lys Pro Ser Arg Ser Glu Met Ser				
	1120	1125	1130	
CAG CAG ACA ACA CCC ATC CCA GCT CCA CGG CCA CCC CTG CCA GTC AAG				3579
Gln Gln Thr Thr Pro Ile Pro Ala Pro Arg Pro Pro Leu Pro Val Lys				
	1135	1140	1145	
AGT CCT GCT GTC CTG CAG CTG CAA CAT TCC AAA GGC AGA GAC TAC CGT				3627
Ser Pro Ala Val Leu Gln Leu Gln His Ser Lys Gly Arg Asp Tyr Arg				
	1150	1155	1160	
GAC AAC ACA GAA CTC CCC CAC CAT GGC AAG CAC CGC CAA GAG GAG GGG				3675
Asp Asn Thr Glu Leu Pro His His Gly Lys His Arg Gln Glu Glu Gly				
	1165	1170	1175	
CTG CTT GGC AGG ACT GCC ATGCAGTGAG CTGCTGGTGA TCGGAGCCTG				3723
Leu Leu Gly Arg Thr Ala				
	1180	1185		
GAGGAACAGC ACAAAGCAGA CCTGCGACCT CTCTCAGGAT GCCTCTCTCA GGATGCCTCT				3783
TGGAGGACCT CCTGCTAGCT CTTCTTGCCCT AGCTTCAAGT CCCAGGCTGT GTATTTTTTT				3843
TCAGGAAACG GCCTCACTTC TCTGTGGTCC AAGAAGTGTG CTGCTGGCTG CCACACTGTG				3903
CGGCAGATGC TAAAGCTGGA TGACAAACGC ACGCCATACA GACAGCAGAC AGCGGCACTG				3963
GGTCTCAGAA CTTGGATTCC TGGGCCTTCT TCCAGTCGCC GTTTTAAAGA AAGGAACTAA				4023
CGGAGCTGCT CATCCGA				4040

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1185 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

- 39 -

Met Val Pro Gly Trp Asn His Gly Asn Ile Thr Arg Ser Lys Ala Glu  
 1 5 10 15  
 Glu Leu Leu Ser Arg Ala Gly Lys Asp Gly Ser Phe Leu Val Arg Ala  
 20 25 30  
 Ser Glu Ser Ile Pro Arg Ala Cys Ala Leu Cys Val Leu Phe Arg Asn  
 35 40 45  
 Cys Val Tyr Thr Tyr Arg Ile Leu Pro Asn Glu Asp Asp Lys Phe Thr  
 50 55 60  
 Val Gln Ala Ser Glu Gly Val Pro Met Arg Phe Phe Thr Lys Leu Asp  
 65 70 75 80  
 Gln Leu Ile Asp Phe Tyr Lys Lys Glu Asn Met Gly Leu Val Thr His  
 85 90 95  
 Leu Gln Tyr Pro Val Pro Leu Glu Glu Glu Asp Ala Ile Asp Glu Ala  
 100 105 110  
 Glu Glu Asp Thr Glu Ser Val Met Ser Pro Pro Glu Leu Pro Pro Arg  
 115 120 125  
 Asn Ile Pro Met Ser Ala Gly Pro Ser Glu Ala Lys Asp Leu Pro Leu  
 130 135 140  
 Ala Thr Glu Asn Pro Arg Ala Pro Glu Val Thr Arg Leu Ser Leu Ser  
 145 150 155 160  
 Glu Thr Leu Phe Gln Arg Leu Gln Ser Met Asp Thr Ser Gly Leu Pro  
 165 170 175  
 Glu Glu His Leu Lys Ala Ile Gln Asp Tyr Leu Ser Thr Gln Leu Leu  
 180 185 190  
 Leu Asp Ser Asp Phe Leu Lys Thr Gly Ser Ser Asn Leu Pro His Leu  
 195 200 205  
 Lys Lys Leu Met Ser Leu Leu Cys Lys Glu Leu His Gly Glu Val Ile  
 210 215 220  
 Arg Thr Leu Pro Ser Leu Glu Ser Leu Gln Arg Leu Phe Asp Gln Gln  
 225 230 235 240  
 Leu Ser Pro Gly Leu Arg Pro Arg Pro Gln Val Pro Gly Glu Ala Ser  
 245 250 255  
 Pro Ile Thr Met Val Ala Lys Leu Ser Gln Leu Thr Ser Leu Leu Ser  
 260 265 270  
 Ser Ile Glu Asp Lys Val Lys Ser Leu Leu His Glu Gly Ser Glu Ser  
 275 280 285  
 Thr Asn Arg Arg Ser Leu Ile Pro Pro Val Thr Phe Glu Val Lys Ser  
 290 295 300  
 Glu Ser Leu Gly Ile Pro Gln Lys Met His Leu Lys Val Asp Val Glu  
 305 310 315 320  
 Ser Gly Lys Leu Ile Val Lys Lys Ser Lys Asp Gly Ser Glu Asp Lys  
 325 330 335  
 Phe Tyr Ser His Lys Lys Ile Leu Gln Leu Ile Lys Ser Gln Lys Phe  
 340 345 350

SUBSTITUTE SHEET (RULE 26)

- 40 -

Leu Asn Lys Leu Val Ile Leu Val Glu Thr Glu Lys Glu Lys Ile Leu  
 355 360 365  
 Arg Lys Glu Tyr Val Phe Ala Asp Ser Lys Lys Arg Glu Gly Phe Cys  
 370 375 380  
 Gln Leu Leu Gln Gln Met Lys Asn Lys His Ser Glu Gln Pro Glu Pro  
 385 390 395 400  
 Asp Met Ile Thr Ile Phe Ile Gly Thr Trp Asn Met Gly Asn Ala Pro  
 405 410 415  
 Pro Pro Lys Lys Ile Thr Ser Trp Phe Leu Ser Lys Gly Gln Gly Lys  
 420 425 430  
 Thr Arg Asp Asp Ser Ala Asp Tyr Ile Pro His Asp Ile Tyr Val Ile  
 435 440 445  
 Gly Thr Gln Glu Asp Pro Leu Gly Glu Lys Glu Trp Leu Glu Leu Leu  
 450 455 460  
 Arg His Ser Leu Gln Glu Val Thr Ser Met Thr Phe Lys Thr Val Ala  
 465 470 475 480  
 Ile His Thr Leu Trp Asn Ile Arg Ile Val Val Leu Ala Lys Pro Glu  
 485 490 495  
 His Glu Asn Arg Ile Ser His Ile Cys Thr Asp Asn Val Lys Thr Gly  
 500 505 510  
 Ile Ala Asn Thr Leu Gly Asn Lys Gly Ala Val Gly Val Ser Phe Met  
 515 520 525  
 Phe Asn Gly Thr Ser Leu Gly Phe Val Asn Ser His Leu Thr Ser Gly  
 530 535 540  
 Ser Glu Lys Lys Leu Arg Arg Asn Gln Asn Tyr Met Asn Ile Leu Arg  
 545 550 555 560  
 Phe Leu Ala Leu Gly Asp Lys Lys Leu Ser Pro Phe Asn Ile Thr His  
 565 570 575  
 Arg Phe Thr His Leu Phe Trp Leu Gly Asp Leu Asn Tyr Arg Val Glu  
 580 585 590  
 Leu Pro Thr Trp Glu Ala Glu Ala Ile Ile Gln Lys Ile Lys Gln Gln  
 595 600 605  
 Gln Tyr Ser Asp Leu Leu Ala His Asp Gln Leu Leu Leu Glu Arg Lys  
 610 615 620  
 Asp Gln Lys Val Phe Leu His Phe Glu Glu Glu Glu Ile Thr Phe Ala  
 625 630 635 640  
 Pro Thr Tyr Arg Phe Glu Arg Leu Thr Arg Asp Lys Tyr Ala Tyr Thr  
 645 650 655  
 Lys Gln Lys Ala Thr Gly Met Lys Tyr Asn Leu Pro Ser Trp Cys Asp  
 660 665 670  
 Arg Val Leu Trp Lys Ser Tyr Pro Leu Val His Val Val Cys Gln Ser  
 675 680 685  
 Tyr Gly Ser Thr Ser Asp Ile Met Thr Ser Asp His Ser Pro Val Phe  
 690 695 700

- 41 -

Ala 705	Thr	Phe	Glu	Ala	Gly 710	Val	Thr	Ser	Gln	Phe 715	Val	Ser	Lys	Asn	Gly 720
Pro	Gly	Thr	Val	Asp 725	Ser	Gln	Gly	Gln	Ile 730	Glu	Phe	Leu	Ala	Cys 735	Tyr
Ala	Thr	Leu	Lys 740	Thr	Lys	Ser	Gln	Thr 745	Lys	Phe	Tyr	Leu	Glu 750	Phe	His
Ser	Ser	Cys 755	Leu	Glu	Ser	Phe	Val 760	Lys	Ser	Gln	Glu	Gly 765	Glu	Asn	Glu
Glu	Gly 770	Ser	Glu	Gly	Glu	Leu 775	Val	Val	Arg	Phe	Gly 780	Glu	Thr	Leu	Pro
Lys 785	Leu	Lys	Pro	Ile	Ile 790	Ser	Asp	Pro	Glu	Tyr 795	Leu	Leu	Asp	Gln	His 800
Ile	Leu	Ile	Ser	Ile 805	Lys	Ser	Ser	Asp	Ser 810	Asp	Glu	Ser	Tyr	Gly 815	Glu
Gly	Cys	Ile	Ala 820	Leu	Arg	Leu	Glu	Thr 825	Thr	Glu	Ala	Gln	His 830	Pro	Ile
Tyr	Thr	Pro 835	Leu	Thr	His	His	Gly 840	Glu	Met	Thr	Gly	His 845	Phe	Arg	Gly
Glu 850	Ile	Lys	Leu	Gln	Thr	Ser 855	Gln	Gly	Lys	Met	Arg 860	Glu	Lys	Leu	Tyr
Asp 865	Phe	Val	Lys	Thr	Glu 870	Arg	Asp	Glu	Ser	Ser 875	Gly	Met	Lys	Cys	Leu 880
Lys	Asn	Leu	Thr	Ser 885	His	Asp	Pro	Met	Arg 890	Gln	Trp	Glu	Pro	Ser 895	Gly
Arg	Val	Pro	Ala 900	Cys	Gly	Val	Ser	Ser 905	Leu	Asn	Glu	Met	Ile 910	Asn	Pro
Asn	Tyr	Ile 915	Gly	Met	Gly	Pro	Phe 920	Gly	Gln	Pro	Leu	His 925	Gly	Lys	Ser
Thr 930	Leu	Ser	Pro	Asp	Gln	Gln 935	Leu	Thr	Ala	Trp	Ser 940	Tyr	Asp	Gln	Leu
Pro 945	Lys	Asp	Ser	Ser	Leu 950	Gly	Pro	Gly	Arg	Gly 955	Glu	Gly	Pro	Pro	Thr 960
Pro	Pro	Ser	Gln	Pro 965	Pro	Leu	Ser	Pro	Lys 970	Lys	Phe	Ser	Ser	Ser 975	Thr
Thr	Asn	Arg	Gly 980	Pro	Cys	Pro	Arg	Val 985	Gln	Glu	Ala	Arg	Pro 990	Gly	Asp
Leu	Gly	Lys 995	Val	Glu	Ala	Leu	Leu 1000	Gln	Glu	Asp	Leu	Leu	Leu 1005	Thr	Lys
Pro 1010	Glu	Met	Phe	Glu	Asn	Pro 1015	Leu	Tyr	Gly	Ser	Val 1020	Ser	Ser	Phe	Pro
Lys 1025	Leu	Val	Pro	Arg	Lys 1030	Glu	Gln	Glu	Ser	Pro 1035	Lys	Met	Leu	Arg	Lys 1040
Glu	Pro	Pro	Pro	Cys 1045	Pro	Asp	Pro	Gly	Ile 1050	Ser	Ser	Pro	Ser	Ile 1055	Val

SUBSTITUTE SHEET (RULE 26)

- 42 -

Leu Pro Lys Ala Gln Glu Val Glu Ser Val Lys Gly Thr Ser Lys Gln  
 1060 1065 1070  
 Ala Pro Val Pro Val Leu Gly Pro Thr Pro Arg Ile Arg Ser Phe Thr  
 1075 1080 1085  
 Cys Ser Ser Ser Ala Glu Gly Arg Met Thr Ser Gly Asp Lys Ser Gln  
 1090 1095 1100  
 Gly Lys Pro Lys Ala Ser Ala Ser Ser Gln Ala Pro Val Pro Val Lys  
 1105 1110 1115 1120  
 Arg Pro Val Lys Pro Ser Arg Ser Glu Met Ser Gln Gln Thr Thr Pro  
 1125 1130 1135  
 Ile Pro Ala Pro Arg Pro Pro Leu Pro Val Lys Ser Pro Ala Val Leu  
 1140 1145 1150  
 Gln Leu Gln His Ser Lys Gly Arg Asp Tyr Arg Asp Asn Thr Glu Leu  
 1155 1160 1165  
 Pro His His Gly Lys His Arg Gln Glu Glu Gly Leu Leu Gly Arg Thr  
 1170 1175 1180  
 Ala  
 1185

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3031 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens
- (B) STRAIN: Shc Proteins

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 82..1503

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GC GGTAACCT AAGCTGGCAG TG GCGTGATC CGGCACCAAA TCGGCCCGCG GTGCGTGCGG	60
AGACTCCATG AGGCCCTGGA C ATG AAC AAG CTG AGT GGA GGC GGC GGG CGC	111
Met Asn Lys Leu Ser Gly Gly Gly Gly Arg	
1 5 10	
AGG ACT CGG GTG GAA GGG GGC CAG CTT GGG GGC GAG GAG TGG ACC CGC	159
Arg Thr Arg Val Glu Gly Gly Gln Leu Gly Gly Glu Glu Trp Thr Arg	
15 20 25	
CAC GGG AGC TTT GTC AAT AAG CCC ACG CGG GGC TGG CTG CAT CCC AAC	207
His Gly Ser Phe Val Asn Lys Pro Thr Arg Gly Trp Leu His Pro Asn	
30 35 40	
GAC AAA GTC ATG GGA CCC GGG GTT TCC TAC TTG GTT CGG TAC ATG GGT	255
Asp Lys Val Met Gly Pro Gly Val Ser Tyr Leu Val Arg Tyr Met Gly	
45 50 55	

TGT Cys	GTG Val	GAG Glu	GTC Val	CTC Leu	CAG Gln	TCA Ser	ATG Met	CGT Arg	GCC Ala	CTG Leu	GAC Asp	TTC Phe	AAC Asn	ACC Thr	CGG Arg	303
60						65					70					
ACT Thr	CAG Gln	GTC Val	ACC Thr	AGG Arg	GAG Glu	GCC Ala	ATC Ile	AGT Ser	CTG Leu	GTG Val	TGT Cys	GAG Glu	GCT Ala	GTG Val	CCG Pro	351
75					80					85					90	
GGT Gly	GCT Ala	AAG Lys	GGG Gly	GCG Ala	ACA Thr	AGG Arg	AGG Arg	AGA Arg	AAG Lys	CCC Pro	TGT Cys	AGC Ser	CGC Arg	CCG Pro	CTC Leu	399
				95					100					105		
AGC Ser	TCT Ser	ATC Ile	CTG Leu	GGG Gly	AGG Arg	AGT Ser	AAC Asn	CTG Leu	AAA Lys	TTT Phe	GCT Ala	GGA Gly	ATG Met	CCA Pro	ATC Ile	447
			110					115					120			
ACT Thr	CTC Leu	ACC Thr	GTC Val	TCC Ser	ACC Thr	AGC Ser	AGC Ser	CTC Leu	AAC Asn	CTC Leu	ATG Met	GCC Ala	GCA Ala	GAC Asp	TGC Cys	495
		125						130				135				
AAA Lys	CAG Gln	ATC Ile	ATC Ile	GCC Ala	AAC Asn	CAC His	CAC His	ATG Met	CAA Gln	TCT Ser	ATC Ile	TCA Ser	TTT Phe	GCA Ala	TCC Ser	543
	140					145					150					
GGC Gly	GGG Gly	GAT Asp	CCG Pro	GAC Asp	ACA Thr	GCC Ala	GAG Glu	TAT Tyr	GTC Val	GCC Ala	TAT Tyr	GTT Val	GCC Ala	AAA Lys	GAC Asp	591
155					160					165					170	
CCT Pro	GTG Val	AAT Asn	CAG Gln	AGA Arg	GCC Ala	TGC Cys	CAC His	ATT Ile	CTG Leu	GAG Glu	TGT Cys	CCC Pro	GAA Glu	GGG Gly	CTT Leu	639
				175					180					185		
GCC Ala	CAG Gln	GAT Asp	GTC Val	ATC Ile	AGC Ser	ACC Thr	ATT Ile	GGC Gly	CAG Gln	GCC Ala	TTC Phe	GAG Glu	TTG Leu	CGC Arg	TTC Phe	687
			190					195					200			
AAA Lys	CAA Gln	TAC Tyr	CTC Leu	AGG Arg	AAC Asn	CCA Pro	CCC Pro	AAA Lys	CTG Leu	GTC Val	ACC Thr	CCT Pro	CAT His	GAC Asp	AGG Arg	735
	205						210					215				
ATG Met	GCT Ala	GGC Gly	TTT Phe	GAT Asp	GGC Gly	TCA Ser	GCA Ala	TGG Trp	GAT Asp	GAG Glu	GAG Glu	GAG Glu	GAA Glu	GAG Glu	CCA Pro	783
	220					225				230						
CCT Pro	GAC Asp	CAT His	CAG Gln	TAC Tyr	TAT Tyr	AAT Asn	GAC Asp	TTC Phe	CCG Pro	GGG Gly	AAG Lys	GAA Glu	CCC Pro	CCC Pro	TTG Leu	831
	235				240					245					250	
GGG Gly	GGG Gly	GTG Val	GTA Val	GAC Asp	ATG Met	AGG Arg	CTT Leu	CGG Arg	GAA Glu	GGA Gly	GCC Ala	GCT Ala	CCA Pro	GGG Gly	GCT Ala	879
				255					260					265		
GCT Ala	CGA Arg	CCC Pro	ACT Thr	GCA Ala	CCC Pro	AAT Asn	GCC Ala	CAG Gln	ACC Thr	CCC Pro	AGC Ser	CAC His	TTG Leu	GGA Gly	GCT Ala	927
			270					275					280			
ACA Thr	TTG Leu	CCT Pro	GTA Val	GGA Gly	CAG Gln	CCT Pro	GTT Val	GGG Gly	GGA Gly	GAT Asp	CCA Pro	GAA Glu	GTC Val	CGC Arg	AAA Lys	975
		285					290					295				
CAG Gln	ATG Met	CCA Pro	CCT Pro	CCA Pro	CCA Pro	CCC Pro	TGT Cys	CCA Pro	GGC Gly	AGA Arg	GAG Glu	CTT Leu	TTT Phe	GAT Asp	GAT Asp	1023
	300					305					310					
CCC Pro	TCC Ser	TAT Tyr	GTC Val	AAC Asn	GTC Val	CAG Gln	AAC Asn	CTA Leu	GAC Asp	AAG Lys	GCC Ala	CGG Arg	CAA Gln	GCA Ala	GTG Val	1071

- 44 -

315	320	325	330	
GGT GGT GCT GGG CCC CCC AAT CCT GCT ATC AAT GGC AGT GCA CCC CGG				1119
Gly Gly Ala Gly Pro Pro Asn Pro Ala Ile Asn Gly Ser Ala Pro Arg				
	335	340	345	
GAC CTG TTT GAC ATG AAG CCC TTC GAA GAT GCT CTT CGG GTG CCT CCA				1167
Asp Leu Phe Asp Met Lys Pro Phe Glu Asp Ala Leu Arg Val Pro Pro				
	350	355	360	
CCT CCC CAG TCG GTG TCC ATG GCT GAG CAG CTC CGA GGG GAG CCC TGG				1215
Pro Pro Gln Ser Val Ser Met Ala Glu Gln Leu Arg Gly Glu Pro Trp				
	365	370	375	
TTC CAT GGG AAG CTG AGC CGG CGG GAG GCT GAG GCA CTG CTG CAG CTC				1263
Phe His Gly Lys Leu Ser Arg Arg Glu Ala Glu Ala Leu Leu Gln Leu				
	380	385	390	
AAT GGG GAC TTC TTG GTA CGG GAG AGC ACG ACC ACA CCT GGC CAG TAT				1311
Asn Gly Asp Phe Leu Val Arg Glu Ser Thr Thr Thr Pro Gly Gln Tyr				
	395	400	405	410
GTG CTC ACT GGC TTG CAG AGT GGG CAG CCT AAG CAT TTG CTA CTG GTG				1359
Val Leu Thr Gly Leu Gln Ser Gly Gln Pro Lys His Leu Leu Leu Val				
	415	420	425	
GAC CCT GAG GGT GTG GTT CGG ACT AAG GAT CAC CGC TTT GAA AGT GTC				1407
Asp Pro Glu Gly Val Val Arg Thr Lys Asp His Arg Phe Glu Ser Val				
	430	435	440	
AGT CAC CTT ATC AGC TAC CAC ATG GAC AAT CAC TTG CCC ATC ATC TCT				1455
Ser His Leu Ile Ser Tyr His Met Asp Asn His Leu Pro Ile Ile Ser				
	445	450	455	
GCG GGC AGC GAA CTG TGT CTA CAG CAA CCT GTG GAG CGG AAA CTG TGA				1503
Ala Gly Ser Glu Leu Cys Leu Gln Gln Pro Val Glu Arg Lys Leu *				
	460	465	470	
TCTGCCCTAG CGCTCTCTTC CAGAAGATGC CCTCCAATCC TTTCCACCCT ATTCCCTAAC				1563
TCTCGGGACC TCGTTTGGGA GTGTTCTGTG GGCTTGGCCT TGTGTCAGAG CTGGGAGTAG				1623
CATGGACTCT GGGTTTCATA TCCAGCTGAG TGAGAGGGTT TGAGTCAAAA GCCTGGGTGA				1683
GAATCCTGCC TCTCCCCAAA CATTAATCAC CAAAGTATTA ATGTACAGAG TGGCCCCCTCA				1743
CCTGGGCCTT TCCTGTGCCA ACCTGATGCC CCTTCCCCAA GAAGGTGAGT GCTTGTCTATG				1803
GAAAATGTCC TGTGGTGACA GGCCAGTGG AACAGTCACC CTTCTGGGCA AGGGGGAACA				1863
AATCACACCT CTGGGCTTCA GGGTATCCCA GACCCCTCTC AACACCCGCC CCCCCATGT				1923
TTAAACTTTG TGCCTTTGAC CATCTCTTAG GTCTAATGAT ATTTTATGCA AACAGTTCTT				1983
GGACCCCTGA ATTCTTCAAT GACAGGGATG CCAACACCTT CTTGGCTTCT GGGACCTGTG				2043
TTCTTGCTGA GCACCCTCTC CGGTTTGGGT TGGGATAACA GAGGCAGGAG TGGCAGCTGT				2103
CCCCTCTCCC TGGGGATATG CAACCCCTTAG AGATTGCCCC AGAGCCCCAC TCCCGGCCAG				2163
GCGGGAGATG GACCCCTCCC TTGCTCAGTG CCTCCTGGCC GGGGCCCCTC ACCCCAAGGG				2223
GTCTGTATAT ACATTTTATA AGGCCTGCCC TCCCATGTTG CATGCCTATG TACTCTGCGC				2283
CAAAGTGCAG CCCTTCCTCC TGAAGCCTCT GCCCTGCCTC CCTTTCTGGG AGGGCGGGGT				2343

- 45 -

```

GGGGGTGACT GAATTTGGGC CTCTTGACATA GTTAACTCTC CCAGGTGGAT TTTGTGGAGG      2403
TGAGAAAAGG GGCATTGAGA CTATAAAGCA GTAGACAATC CCCACATACC ATCTGTAGAG      2463
TTGGAACTGC ATTCTTTTAA AGTTTTATAT GCATATATTT TAGGGCTGCT AGACTTACTT      2523
TCCTATTTTC TTTTCCATTG CTTATTCTTG AGCACAAAAT GATAATCAAT TATTACATTT      2583
ATACATCACC TTTTGTACTT TTCCAAGCCC TTTTACAGCT CTTGGCATTT TCCTCGCCTA      2643
GGCCTGTGAG GTAAC TGGGA TCGCACCTTT TATACCAGAG ACCTGAGGCA GATGAAATTT      2703
ATTTCCATCT AGGACTAGAA AAAC TTGGGT CTCTTACCGC GAGACTGAGA GGCAGAAGTC      2763
AGCCCGAATG CCTGTCAGTT TCATGGAGGG GAAACGCAAA ACCTGCAGTT CCTGAGTACC      2823
TTCTACAGGC CCGGCCCCAGC CTAGGCCCCG GGTGGCCACA CCACAGCAAG CCGGCCCCCC      2883
CTCTTTTGGC CTTGTGGATA AGGGAGAGTT GACCGTTTTT ATCCTGGCCT CTTTTTGCTG      2943
TTTGGATGTT TCCACGGGTC TCACTTATAC CAAAGGGAAA ACTCTTCATT AAAGTCCCGT      3003
ATTTCTTCTA AAAAAAAAAA AAAAAAAA      3031

```

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 474 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Asn Lys Leu Ser Gly Gly Gly Gly Arg Arg Thr Arg Val Glu Gly
 1           5           10           15
Gly Gln Leu Gly Gly Glu Glu Trp Thr Arg His Gly Ser Phe Val Asn
          20           25           30
Lys Pro Thr Arg Gly Trp Leu His Pro Asn Asp Lys Val Met Gly Pro
          35           40           45
Gly Val Ser Tyr Leu Val Arg Tyr Met Gly Cys Val Glu Val Leu Gln
          50           55           60
Ser Met Arg Ala Leu Asp Phe Asn Thr Arg Thr Gln Val Thr Arg Glu
          65           70           75           80
Ala Ile Ser Leu Val Cys Glu Ala Val Pro Gly Ala Lys Gly Ala Thr
          85           90           95
Arg Arg Arg Lys Pro Cys Ser Arg Pro Leu Ser Ser Ile Leu Gly Arg
          100          105          110
Ser Asn Leu Lys Phe Ala Gly Met Pro Ile Thr Leu Thr Val Ser Thr
          115          120          125
Ser Ser Leu Asn Leu Met Ala Ala Asp Cys Lys Gln Ile Ile Ala Asn
          130          135          140
His His Met Gln Ser Ile Ser Phe Ala Ser Gly Gly Asp Pro Asp Thr
          145          150          155          160

```

- 46 -

Ala Glu Tyr Val Ala Tyr Val Ala Lys Asp Pro Val Asn Gln Arg Ala  
 165 170 175  
 Cys His Ile Leu Glu Cys Pro Glu Gly Leu Ala Gln Asp Val Ile Ser  
 180 185 190  
 Thr Ile Gly Gln Ala Phe Glu Leu Arg Phe Lys Gln Tyr Leu Arg Asn  
 195 200 205  
 Pro Pro Lys Leu Val Thr Pro His Asp Arg Met Ala Gly Phe Asp Gly  
 210 215 220  
 Ser Ala Trp Asp Glu Glu Glu Glu Glu Pro Pro Asp His Gln Tyr Tyr  
 225 230 235 240  
 Asn Asp Phe Pro Gly Lys Glu Pro Pro Leu Gly Gly Val Val Asp Met  
 245 250 255  
 Arg Leu Arg Glu Gly Ala Ala Pro Gly Ala Ala Arg Pro Thr Ala Pro  
 260 265 270  
 Asn Ala Gln Thr Pro Ser His Leu Gly Ala Thr Leu Pro Val Gly Gln  
 275 280 285  
 Pro Val Gly Gly Asp Pro Glu Val Arg Lys Gln Met Pro Pro Pro Pro  
 290 295 300  
 Pro Cys Pro Gly Arg Glu Leu Phe Asp Asp Pro Ser Tyr Val Asn Val  
 305 310 315 320  
 Gln Asn Leu Asp Lys Ala Arg Gln Ala Val Gly Gly Ala Gly Pro Pro  
 325 330 335  
 Asn Pro Ala Ile Asn Gly Ser Ala Pro Arg Asp Leu Phe Asp Met Lys  
 340 345 350  
 Pro Phe Glu Asp Ala Leu Arg Val Pro Pro Pro Pro Gln Ser Val Ser  
 355 360 365  
 Met Ala Glu Gln Leu Arg Gly Glu Pro Trp Phe His Gly Lys Leu Ser  
 370 375 380  
 Arg Arg Glu Ala Glu Ala Leu Leu Gln Leu Asn Gly Asp Phe Leu Val  
 385 390 395 400  
 Arg Glu Ser Thr Thr Thr Pro Gly Gln Tyr Val Leu Thr Gly Leu Gln  
 405 410 415  
 Ser Gly Gln Pro Lys His Leu Leu Leu Val Asp Pro Glu Gly Val Val  
 420 425 430  
 Arg Thr Lys Asp His Arg Phe Glu Ser Val Ser His Leu Ile Ser Tyr  
 435 440 445  
 His Met Asp Asn His Leu Pro Ile Ile Ser Ala Gly Ser Glu Leu Cys  
 450 455 460  
 Leu Gln Gln Pro Val Glu Arg Lys Leu \*  
 465 470

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1109 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single

SUBSTITUTE SHEET (RULE 26)

- 47 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(B) STRAIN: GRB2

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 79..732

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```

GCCAGTGAAT TCGGGGGCTC AGCCCTCCTC CCTCCCTTCC CCCTGCTTCA GGCTGCTGAG      60
CACTGAGCAG CGCTCAGA ATG GAA GCC ATC GCC AAA TAT GAC TTC AAA GCT      111
               Met Glu Ala Ile Ala Lys Tyr Asp Phe Lys Ala
                   1                   5                   10
ACT GCA GAC GAC GAG CTG AGC TTC AAA AGG GGG GAC ATC CTC AAG GTT      159
Thr Ala Asp Asp Glu Leu Ser Phe Lys Arg Gly Asp Ile Leu Lys Val
                   15                   20                   25
TTG AAC GAA GAA TGT GAT CAG AAC TGG TAC AAG GCA GAG CTT AAT GGA      207
Leu Asn Glu Glu Cys Asp Gln Asn Trp Tyr Lys Ala Glu Leu Asn Gly
                   30                   35                   40
AAA GAC GGC TTC ATT CCC AAG AAC TAC ATA GAA ATG AAA CCA CAT CCG      255
Lys Asp Gly Phe Ile Pro Lys Asn Tyr Ile Glu Met Lys Pro His Pro
                   45                   50                   55
TGG TTT TTT GGC AAA ATC CCC AGA GCC AAG GCA GAA GAA ATG CTT AGC      303
Trp Phe Phe Gly Lys Ile Pro Arg Ala Lys Ala Glu Glu Met Leu Ser
                   60                   65                   70                   75
AAA CAG CGG CAC GAT GGG GCC TTT CTT ATC CGA GAG AGT GAG AGC GCT      351
Lys Gln Arg His Asp Gly Ala Phe Leu Ile Arg Glu Ser Glu Ser Ala
                   80                   85                   90
CCT GGG GAC TTC TCC CTC TCT GTC AAG TTT GGA AAC GAT GTG CAG CAC      399
Pro Gly Asp Phe Ser Leu Ser Val Lys Phe Gly Asn Asp Val Gln His
                   95                   100                   105
TTC AAG GTG CTC CGA GAT GGA GCC GGG AAG TAC TTC CTC TGG GTG GTG      447
Phe Lys Val Leu Arg Asp Gly Ala Gly Lys Tyr Phe Leu Trp Val Val
                   110                   115                   120
AAG TTC AAT TCT TTG AAT GAG CTG GTG GAT TAT CAC AGA TCT ACA TCT      495
Lys Phe Asn Ser Leu Asn Glu Leu Val Asp Tyr His Arg Ser Thr Ser
                   125                   130                   135
GTC TCC AGA AAC CAG CAG ATA TTC CTG CGG GAC ATA GAA CAG GTG CCA      543
Val Ser Arg Asn Gln Gln Ile Phe Leu Arg Asp Ile Glu Gln Val Pro
                   140                   145                   150                   155
CAG CAG CCG ACA TAC GTC CAG GCC CTC TTT GAC TTT GAT CCC CAG GAG      591
Gln Gln Pro Thr Tyr Val Gln Ala Leu Phe Asp Phe Asp Pro Gln Glu
                   160                   165                   170
GAT GGA GAG CTG GGC TTC CGC CGG GGA GAT TTT ATC CAT GTC ATG GAT      639
Asp Gly Glu Leu Gly Phe Arg Arg Gly Asp Phe Ile His Val Met Asp
                   175                   180                   185
AAC TCA GAC CCC AAC TGG TGG AAA GGA GCT TGC CAC GGG CAG ACC GGC      687

```

SUBSTITUTE SHEET (RULE 26)

- 48 -

Asn	Ser	Asp	Pro	Asn	Trp	Trp	Lys	Gly	Ala	Cys	His	Gly	Gln	Thr	Gly		
		190					195					200					
ATG	TTT	CCC	CGC	AAT	TAT	GTC	ACC	CCC	GTG	AAC	CGG	AAC	GTC	TAA			732
Met	Phe	Pro	Arg	Asn	Tyr	Val	Thr	Pro	Val	Asn	Arg	Asn	Val	*			
	205					210					215						
GAGTCAAGAA	GCAATTATTT	AAAGAAAGTG	AAAAATGTAA	AACACATACA	AAAGAATTAA												792
ACCCACAAGC	TGCCTCTGAC	AGCAGCCTGT	GAGGGAGTGC	AGAACACCTG	GCCGGGTCAC												852
CCTGTGACCC	TCTCACTTTG	GTTGGAACCT	TAGGGGGTGG	GAGGGGGCGT	TGGATTTAAA												912
AATGCCAAAA	CTTACCTATA	AATTAAGAAG	AGTTTTTATT	ACAAATTTTC	ACTGCTGCTC												972
CTCTTTCCCC	TCCTTTGTCT	TTTTTTTCAT	CCTTTTTTCT	CTTCTGTCCA	TCAGTGCATG												1032
ACGTTTAAGG	CCACGTATAG	TCCTAGCTGA	CGCCAATAAT	AAAAACAAG	AAACCAAAAA												1092
AAAAAAACCC	GAATTCA																1109

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 218 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met	Glu	Ala	Ile	Ala	Lys	Tyr	Asp	Phe	Lys	Ala	Thr	Ala	Asp	Asp	Glu		
1				5					10					15			
Leu	Ser	Phe	Lys	Arg	Gly	Asp	Ile	Leu	Lys	Val	Leu	Asn	Glu	Glu	Cys		
			20					25					30				
Asp	Gln	Asn	Trp	Tyr	Lys	Ala	Glu	Leu	Asn	Gly	Lys	Asp	Gly	Phe	Ile		
		35				40						45					
Pro	Lys	Asn	Tyr	Ile	Glu	Met	Lys	Pro	His	Pro	Trp	Phe	Phe	Gly	Lys		
		50				55					60						
Ile	Pro	Arg	Ala	Lys	Ala	Glu	Glu	Met	Leu	Ser	Lys	Gln	Arg	His	Asp		
65				70					75					80			
Gly	Ala	Phe	Leu	Ile	Arg	Glu	Ser	Glu	Ser	Ala	Pro	Gly	Asp	Phe	Ser		
			85					90					95				
Leu	Ser	Val	Lys	Phe	Gly	Asn	Asp	Val	Gln	His	Phe	Lys	Val	Leu	Arg		
			100					105					110				
Asp	Gly	Ala	Gly	Lys	Tyr	Phe	Leu	Trp	Val	Val	Lys	Phe	Asn	Ser	Leu		
		115					120					125					
Asn	Glu	Leu	Val	Asp	Tyr	His	Arg	Ser	Thr	Ser	Val	Ser	Arg	Asn	Gln		
		130				135						140					
Gln	Ile	Phe	Leu	Arg	Asp	Ile	Glu	Gln	Val	Pro	Gln	Gln	Pro	Thr	Tyr		
145				150						155					160		
Val	Gln	Ala	Leu	Phe	Asp	Phe	Asp	Pro	Gln	Glu	Asp	Gly	Glu	Leu	Gly		
			165					170						175			

SUBSTITUTE SHEET (RULE 26)

- 49 -

Phe Arg Arg Gly Asp Phe Ile His Val Met Asp Asn Ser Asp Pro Asn  
 180 185 190  
 Trp Trp Lys Gly Ala Cys His Gly Gln Thr Gly Met Phe Pro Arg Asn  
 195 200 205  
 Tyr Val Thr Pro Val Asn Arg Asn Val \*  
 210 215

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4870 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: hSHIP

## (ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 113..3673

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CCCAAGAGGC AACGGGCGGC AGGTTGCACT GGAGGGGCCT CCGCTCCCCT CGGTGGTGTG 60  
 TGGGTCCTGG GGGTGCCTGC CGGCCAGCC GAGGAGGCC ACGCCACCA TG GTC 115  
 Val  
 1  
 CCC TGC TGG AAC CAT GGC AAC ATC ACC CGC TCC AAG GCG GAG GAG CTG 163  
 Pro Cys Trp Asn His Gly Asn Ile Thr Arg Ser Lys Ala Glu Glu Leu  
 5 10 15  
 CTT TGC AGG ACA GGC AAG GAC GGG AGC TTC CTC GTG CGT GCC AGC GAG 211  
 Leu Cys Arg Thr Gly Lys Asp Gly Ser Phe Leu Val Arg Ala Ser Glu  
 20 25 30  
 TCC ATC TTC CGG GCA TAC GCG CTC TGC GTG CTG TAT CGG AAT TGC GTT 259  
 Ser Ile Phe Arg Ala Tyr Ala Leu Cys Val Leu Tyr Arg Asn Cys Val  
 35 40 45  
 TAT ACT TAC AGA ATT CTG CCC AAT GAA GAT GAT AAA TTC ACT GTT CAG 307  
 Tyr Thr Tyr Arg Ile Leu Pro Asn Glu Asp Asp Lys Phe Thr Val Gln  
 50 55 60 65  
 GCA TCC GAA GGC GTC TCC ATG AGG TTC TTC ACC AAG CTG GAC CAG CTC 355  
 Ala Ser Glu Gly Val Ser Met Arg Phe Phe Thr Lys Leu Asp Gln Leu  
 70 75 80  
 ATC GAG TTT TAC AAG AAG GAA AAC ATG GGG CTG GTG ACC CAT CTG CAA 403  
 Ile Glu Phe Tyr Lys Lys Glu Asn Met Gly Leu Val Thr His Leu Gln  
 85 90 95  
 TAC CCT GTG CCG CTG GAG GAA GAG GAC ACA GGC GAC GAC CCT GAG GAG 451  
 Tyr Pro Val Pro Leu Glu Glu Glu Asp Thr Gly Asp Asp Pro Glu Glu  
 100 105 110

SUBSTITUTE SHEET (RULE 26)

GAC Asp	ACA Thr	GAA Glu	AGT Ser	GTC Val	GTG Val	TCT Ser	CCA Pro	CCC Pro	GAG Glu	CTG Leu	CCC Pro	CCA Pro	AGA Arg	AAC Asn	ATC Ile	499
115						120					125					
CCG Pro	CTG Leu	ACT Thr	GCC Ala	AGC Ser	TCC Ser	TGT Cys	GAG Glu	GCC Ala	AAG Lys	GAG Glu	GTT Val	CCT Pro	TTT Phe	TCA Ser	AAC Asn	547
130					135					140					145	
GAG Glu	AAT Asn	CCC Pro	CGA Arg	GCG Ala	ACC Thr	GAG Glu	ACC Thr	AGC Ser	CGG Arg	CCG Pro	AGC Ser	CTC Leu	TCC Ser	GAG Glu	ACA Thr	595
				150					155						160	
TTG Leu	TTC Phe	CAG Gln	CGA Arg	CTG Leu	CAA Gln	AGC Ser	ATG Met	GAC Asp	ACC Thr	AGT Ser	GGG Gly	CTT Leu	CCA Pro	GAA Glu	GAG Glu	643
			165					170					175			
CAT His	CTT Leu	AAG Lys	GCC Ala	ATC Ile	CAA Gln	GAT Asp	TAT Tyr	TTA Leu	AGC Ser	ACT Thr	CAG Gln	CTC Leu	GCC Ala	CAG Gln	GAC Asp	691
		180					185					190				
TCT Ser	GAA Glu	TTT Phe	GTG Val	AAG Lys	ACA Thr	GGG Gly	TCC Ser	AGC Ser	AGT Ser	CTT Leu	CCT Pro	CAC His	CTG Leu	AAG Lys	AAA Lys	739
	195					200					205					
CTG Leu	ACC Thr	ACA Thr	CTG Leu	CTC Leu	TGC Cys	AAG Lys	GAG Glu	CTC Leu	TAT Tyr	GGA Gly	GAA Glu	GTC Val	ATC Ile	CGG Arg	ACC Thr	787
210					215					220					225	
CTC Leu	CCA Pro	TCC Ser	CTG Leu	GAG Glu	TCT Ser	CTG Leu	CAG Gln	AGG Arg	TTA Leu	TTT Phe	GAC Asp	CAG Gln	CAG Gln	CTC Leu	TCC Ser	835
				230					235					240		
CCG Pro	GGC Gly	CTC Leu	CGT Arg	CCA Pro	CGT Arg	CCT Pro	CAG Gln	GTT Val	CCT Pro	GGT Gly	GAG Glu	GCC Ala	AAT Asn	CCC Pro	ATC Ile	883
			245					250					255			
AAC Asn	ATG Met	GTG Val	TCC Ser	AAG Lys	CTC Leu	AGC Ser	CAA Gln	CTG Leu	ACA Thr	AGC Ser	CTG Leu	TTG Leu	TCA Ser	TCC Ser	ATT Ile	931
		260					265					270				
GAA Glu	GAC Asp	AAG Lys	GTC Val	AAG Lys	GCC Ala	TTG Leu	CTG Leu	CAC His	GAG Glu	GGT Gly	CCT Pro	GAG Glu	TCT Ser	CCG Pro	CAC His	979
	275					280					285					
CGG Arg	CCC Pro	TCC Ser	CTT Leu	ATC Ile	CCT Pro	CCA Pro	GTC Val	ACC Thr	TTT Phe	GAG Glu	GTG Val	AAG Lys	GCA Ala	GAG Glu	TCT Ser	1027
290					295					300					305	
CTG Leu	GGG Gly	ATT Ile	CCT Pro	CAG Gln	AAA Lys	ATG Met	CAG Gln	CTC Leu	AAA Lys	GTC Val	GAC Asp	GTT Val	GAG Glu	TCT Ser	GGG Gly	1075
			310						315					320		
AAA Lys	CTG Leu	ATC Ile	ATT Ile	AAG Lys	AAG Lys	TCC Ser	AAG Lys	GAT Asp	GGT Gly	TCT Ser	GAG Glu	GAC Asp	AAG Lys	TTC Phe	TAC Tyr	1123
			325					330					335			
AGC Ser	CAC His	AAG Lys	AAA Lys	ATC Ile	CTG Leu	CAG Gln	CTC Leu	ATT Ile	AAG Lys	TCA Ser	CAG Gln	AAA Lys	TTT Phe	CTG Leu	AAT Asn	1171
		340					345					350				
AAG Lys	TTG Leu	GTG Val	ATC Ile	TTG Leu	GTG Val	GAA Glu	ACA Thr	GAG Glu	AAG Lys	GAG Glu	AAG Lys	ATC Ile	CTG Leu	CGG Arg	AAG Lys	1219
		355				360					365					
GAA Glu	TAT Tyr	GTT Val	TTT Phe	GCT Ala	GAC Asp	TCC Ser	AAA Lys	AAG Lys	AGA Arg	GAA Glu	GGC Gly	TTC Phe	TGC Cys	CAG Gln	CTC Leu	1267

370					375					380					385	
CTG	CAG	CAG	ATG	AAG	AAC	AAG	CAC	TCA	GAG	CAG	CCG	GAG	CCC	GAC	ATG	1315
Leu	Gln	Gln	Met	Lys	Asn	Lys	His	Ser	Glu	Gln	Pro	Glu	Pro	Asp	Met	
				390					395					400		
ATC	ACC	ATC	TTC	ATC	GGC	ACC	TGG	AAC	ATG	GGT	AAC	GCC	CCC	CCT	CCC	1363
Ile	Thr	Ile	Phe	Ile	Gly	Thr	Trp	Asn	Met	Gly	Asn	Ala	Pro	Pro	Pro	
			405					410					415			
AAG	AAG	ATC	ACG	TCC	TGG	TTT	CTC	TCC	AAG	GGG	CAG	GGA	AAG	ACG	CGG	1411
Lys	Lys	Ile	Thr	Ser	Trp	Phe	Leu	Ser	Lys	Gly	Gln	Gly	Lys	Thr	Arg	
		420					425					430				
GAC	GAC	TCT	GCG	GAC	TAC	ATC	CCC	CAT	GAC	ATT	TAC	GTG	ATC	GGC	ACC	1459
Asp	Asp	Ser	Ala	Asp	Tyr	Ile	Pro	His	Asp	Ile	Tyr	Val	Ile	Gly	Thr	
	435					440					445					
CAA	GAG	GAC	CCC	CTG	AGT	GAG	AAG	GAG	TGG	CTG	GAG	ATC	CTC	AAA	CAC	1507
Gln	Glu	Asp	Pro	Leu	Ser	Glu	Lys	Glu	Trp	Leu	Glu	Ile	Leu	Lys	His	
450					455					460					465	
TCC	CTG	CAA	GAA	ATC	ACC	AGT	GTG	ACT	TTT	AAA	ACA	GTC	GCC	ATC	CAC	1555
Ser	Leu	Gln	Glu	Ile	Thr	Ser	Val	Thr	Phe	Lys	Thr	Val	Ala	Ile	His	
				470					475					480		
ACG	CTC	TGG	AAC	ATC	CGC	ATC	GTG	GTG	CTG	GCC	AAG	CCT	GAG	CAC	GAG	1603
Thr	Leu	Trp	Asn	Ile	Arg	Ile	Val	Val	Leu	Ala	Lys	Pro	Glu	His	Glu	
			485				490						495			
AAC	CGG	ATC	AGC	CAC	ATC	TGT	ACT	GAC	AAC	GTG	AAG	ACA	GGC	ATT	GCA	1651
Asn	Arg	Ile	Ser	His	Ile	Cys	Thr	Asp	Asn	Val	Lys	Thr	Gly	Ile	Ala	
	500					505						510				
AAC	ACA	CTG	GGG	AAC	AAG	GGA	GCC	GTG	GGG	GTG	TCG	TTC	ATG	TTC	AAT	1699
Asn	Thr	Leu	Gly	Asn	Lys	Gly	Ala	Val	Gly	Val	Ser	Phe	Met	Phe	Asn	
	515					520					525					
GGA	ACC	TCC	TTA	GGG	TTC	GTC	AAC	AGC	CAC	TTG	ACT	TCA	GGA	AGT	GAA	1747
Gly	Thr	Ser	Leu	Gly	Phe	Val	Asn	Ser	His	Leu	Thr	Ser	Gly	Ser	Glu	
530					535					540					545	
AAG	AAA	CTC	AGG	CGA	AAC	CAA	AAC	TAT	ATG	AAC	ATT	CTC	CGG	TTC	CTG	1795
Lys	Lys	Leu	Arg	Arg	Asn	Gln	Asn	Tyr	Met	Asn	Ile	Leu	Arg	Phe	Leu	
			550						555					560		
GCC	CTG	GGC	GAC	AAG	AAG	CTG	AGT	CCC	TTT	AAC	ATC	ACT	CAC	CGC	TTC	1843
Ala	Leu	Gly	Asp	Lys	Lys	Leu	Ser	Pro	Phe	Asn	Ile	Thr	His	Arg	Phe	
			565					570					575			
ACG	CAC	CTC	TTC	TGG	TTT	GGG	GAT	CTT	AAC	TAC	CGT	GTG	GAT	CTG	CCT	1891
Thr	His	Leu	Phe	Trp	Phe	Gly	Asp	Leu	Asn	Tyr	Arg	Val	Asp	Leu	Pro	
		580					585					590				
ACC	TGG	GAG	GCA	GAA	ACC	ATC	ATC	CAA	AAA	ATC	AAG	CAG	CAG	CAG	TAC	1939
Thr	Trp	Glu	Ala	Glu	Thr	Ile	Ile	Gln	Lys	Ile	Lys	Gln	Gln	Gln	Tyr	
	595					600					605					
GCA	GAC	CTC	CTG	TCC	CAC	GAC	CAG	CTG	CTC	ACA	GAG	AGG	AGG	GAG	CAG	1987
Ala	Asp	Leu	Leu	Ser	His	Asp	Gln	Leu	Leu	Thr	Glu	Arg	Arg	Glu	Gln	
610					615					620					625	
AAG	GTC	TTC	CTA	CAC	TTC	GAG	GAG	GAA	GAA	ATC	ACG	TTT	GCC	CCA	ACC	2035
Lys	Val	Phe	Leu	His	Phe	Glu	Glu	Glu	Glu	Ile	Thr	Phe	Ala	Pro	Thr	
				630				635						640		

TAC	CGT	TTT	GAG	AGA	CTG	ACT	CGG	GAC	AAA	TAC	GCC	TAC	ACC	AAG	CAG	2083
Tyr	Arg	Phe	Glu	Arg	Leu	Thr	Arg	Asp	Lys	Tyr	Ala	Tyr	Thr	Lys	Gln	
			645					650					655			
AAA	GCG	ACA	GGG	ATG	AAG	TAC	AAC	TTG	CCT	TCC	TGG	TGT	GAC	CGA	GTC	2131
Lys	Ala	Thr	Gly	Met	Lys	Tyr	Asn	Leu	Pro	Ser	Trp	Cys	Asp	Arg	Val	
		660					665					670				
CTC	TGG	AAG	TCT	TAT	CCC	CTG	GTG	CAC	GTG	GTG	TGT	CAG	TCT	TAT	GGC	2179
Leu	Trp	Lys	Ser	Tyr	Pro	Leu	Val	His	Val	Val	Cys	Gln	Ser	Tyr	Gly	
	675					680					685					
AGT	ACC	AGC	GAC	ATC	ATG	ACG	AGT	GAC	CAC	AGC	CCT	GTC	TTT	GCC	ACA	2227
Ser	Thr	Ser	Asp	Ile	Met	Thr	Ser	Asp	His	Ser	Pro	Val	Phe	Ala	Thr	
690					695					700					705	
TTT	GAG	GCA	GGA	GTC	ACT	TCC	CAG	TTT	GTC	TCC	AAG	AAC	GGT	CCC	GGG	2275
Phe	Glu	Ala	Gly	Val	Thr	Ser	Gln	Phe	Val	Ser	Lys	Asn	Gly	Pro	Gly	
			710					715					720			
ACT	GTT	GAC	AGC	CAA	GGA	CAG	ATT	GAG	TTT	CTC	AGG	TGC	TAT	GCC	ACA	2323
Thr	Val	Asp	Ser	Gln	Gly	Gln	Ile	Glu	Phe	Leu	Arg	Cys	Tyr	Ala	Thr	
			725					730					735			
TTG	AAG	ACC	AAG	TCC	CAG	ACC	AAA	TTC	TAC	CTG	GAG	TTC	CAC	TCG	AGC	2371
Leu	Lys	Thr	Lys	Ser	Gln	Thr	Lys	Phe	Tyr	Leu	Glu	Phe	His	Ser	Ser	
		740					745					750				
TGC	TTG	GAG	AGT	TTT	GTC	AAG	AGT	CAG	GAA	GGA	GAA	AAT	GAA	GAA	GGA	2419
Cys	Leu	Glu	Ser	Phe	Val	Lys	Ser	Gln	Glu	Gly	Glu	Asn	Glu	Glu	Gly	
	755					760					765					
AGT	GAG	GGG	GAG	CTG	GTG	GTG	AAG	TTT	GGT	GAG	ACT	CTT	CCA	AAG	CTG	2467
Ser	Glu	Gly	Glu	Leu	Val	Val	Lys	Phe	Gly	Glu	Thr	Leu	Pro	Lys	Leu	
770					775					780					785	
AAG	CCC	ATT	ATC	TCT	GAC	CCT	GAG	TAC	CTG	CTA	GAC	CAG	CAC	ATC	CTC	2515
Lys	Pro	Ile	Ile	Ser	Asp	Pro	Glu	Tyr	Leu	Leu	Asp	Gln	His	Ile	Leu	
			790					795						800		
ATC	AGC	ATC	AAG	TCC	TCT	GAC	AGC	GAC	GAA	TCC	TAT	GGC	GAG	GGC	TGC	2563
Ile	Ser	Ile	Lys	Ser	Ser	Asp	Ser	Asp	Glu	Ser	Tyr	Gly	Glu	Gly	Cys	
			805					810					815			
ATT	GCC	CTT	CGG	TTA	GAG	GCC	ACA	GAA	ACG	CAG	CTG	CCC	ATC	TAC	ACG	2611
Ile	Ala	Leu	Arg	Leu	Glu	Ala	Thr	Glu	Thr	Gln	Leu	Pro	Ile	Tyr	Thr	
		820					825					830				
CCT	CTC	ACC	CAC	CAT	GGG	GAG	TTG	ACA	GGC	CAC	TTC	CAG	GGG	GAG	ATC	2659
Pro	Leu	Thr	His	His	Gly	Glu	Leu	Thr	Gly	His	Phe	Gln	Gly	Glu	Ile	
	835					840					845					
AAG	CTG	CAG	ACC	TCT	CAG	GGC	AAG	ACG	AGG	GAG	AAG	CTC	TAT	GAC	TTT	2707
Lys	Leu	Gln	Thr	Ser	Gln	Gly	Lys	Thr	Arg	Glu	Lys	Leu	Tyr	Asp	Phe	
850					855					860					865	
GTG	AAG	ACG	GAG	CGT	GAT	GAA	TCC	AGT	GGG	CCA	AAG	ACC	CTG	AAG	AGC	2755
Val	Lys	Thr	Glu	Arg	Asp	Glu	Ser	Ser	Gly	Pro	Lys	Thr	Leu	Lys	Ser	
			870						875					880		
CTC	ACC	AGC	CAC	GAC	CCC	ATG	AAG	CAG	TGG	GAA	GTC	ACT	AGC	AGG	GCC	2803
Leu	Thr	Ser	His	Asp	Pro	Met	Lys	Gln	Trp	Glu	Val	Thr	Ser	Arg	Ala	
			885					890					895			
CCT	CCG	TGC	AGT	GGC	TCC	AGC	ATC	ACT	GAA	ATC	ATC	AAC	CCC	AAC	TAC	2851
Pro	Pro	Cys	Ser	Gly	Ser	Ser	Ile	Thr	Glu	Ile	Ile	Asn	Pro	Asn	Tyr	

900	905	910	
ATG GGA GTG GGG CCC TTT GGG CCA CCA ATG CCC CTG CAC GTG AAG CAG Met Gly Val Gly Pro Phe Gly Pro Pro Met Pro Leu His Val Lys Gln 915 920 925			2899
ACC TTG TCC CCT GAC CAG CAG CCC ACA GCC TGG AGC TAC GAC CAG CCG Thr Leu Ser Pro Asp Gln Gln Pro Thr Ala Trp Ser Tyr Asp Gln Pro 930 935 940 945			2947
CCC AAG GAC TCC CCG CTG GGG CCC TGC AGG GGA GAA AGT CCT CCG ACA Pro Lys Asp Ser Pro Leu Gly Pro Cys Arg Gly Glu Ser Pro Pro Thr 950 955 960			2995
CCT CCC GGC CAG CCG CCC ATA TCA CCC AAG AAG TTT TTA CCC TCA ACA Pro Pro Gly Gln Pro Pro Ile Ser Pro Lys Lys Phe Leu Pro Ser Thr 965 970 975			3043
GCA AAC CGG GGT CTC CCT CCC AGG ACA CAG GAG TCA AGG CCC AGT GAC Ala Asn Arg Gly Leu Pro Pro Arg Thr Gln Glu Ser Arg Pro Ser Asp 980 985 990			3091
CTG GGG AAG AAC GCA GGG GAC ACG CTG CCT CAG GAG GAC CTG CCG CTG Leu Gly Lys Asn Ala Gly Asp Thr Leu Pro Gln Glu Asp Leu Pro Leu 995 1000 1005			3139
ACG AAG CCC GAG ATG TTT GAG AAC CCC CTG TAT GGG TCC CTG AGT TCC Thr Lys Pro Glu Met Phe Glu Asn Pro Leu Tyr Gly Ser Leu Ser Ser 1010 1015 1020 1025			3187
TTC CCT AAG CCT GCT CCC AGG AAG GAC CAG GAA TCC CCC AAA ATG CCG Phe Pro Lys Pro Ala Pro Arg Lys Asp Gln Glu Ser Pro Lys Met Pro 1030 1035 1040			3235
CGG AAG GAA CCC CCG CCC TGC CCG GAA CCC GGC ATC TTG TCG CCC AGC Arg Lys Glu Pro Pro Pro Cys Pro Glu Pro Gly Ile Leu Ser Pro Ser 1045 1050 1055			3283
ATC GTG CTC ACC AAA GCC CAG GAG GCT GAT CGC GGC GAG GGG CCC GGC Ile Val Leu Thr Lys Ala Gln Glu Ala Asp Arg Gly Glu Gly Pro Gly 1060 1065 1070			3331
AAG CAG GTG CCC GCG CCC CGG CTG CGC TCC TTC ACG TGC TCA TCC TCT Lys Gln Val Pro Ala Pro Arg Leu Arg Ser Phe Thr Cys Ser Ser Ser 1075 1080 1085			3379
GCC GAG GGC AGG GCG GCC GGC GGG GAC AAG AGC CAA GGG AAG CCC AAG Ala Glu Gly Arg Ala Ala Gly Gly Asp Lys Ser Gln Gly Lys Pro Lys 1090 1095 1100 1105			3427
ACC CCG GTC AGC TCC CAG GCC CCG GTG CCG GCC AAG AGG CCC ATC AAG Thr Pro Val Ser Ser Gln Ala Pro Val Pro Ala Lys Arg Pro Ile Lys 1110 1115 1120			3475
CCT TCC AGA TCG GAA ATC AAC CAG CAG ACC CCG CCC ACC CCG ACG CCG Pro Ser Arg Ser Glu Ile Asn Gln Gln Thr Pro Pro Thr Pro Thr Pro 1125 1130 1135			3523
CGG CCG CCG CTG CCA GTC AAG AGC CCG GCG GTG CTG CAC CTC CAG CAC Arg Pro Pro Leu Pro Val Lys Ser Pro Ala Val Leu His Leu Gln His 1140 1145 1150			3571
TCC AAG GGC CGC GAC TAC CGC GAC AAC ACC GAG CTC CCG CAT CAC GGC Ser Lys Gly Arg Asp Tyr Arg Asp Asn Thr Glu Leu Pro His His Gly 1155 1160 1165			3619

- 54 -

AAG CAC CGG CCG GAG GAG GGG CCA CCA GGG CCT CTA GGC AGG ACT GCC 3667  
 Lys His Arg Pro Glu Glu Gly Pro Pro Gly Pro Leu Gly Arg Thr Ala  
 1170 1175 1180 1185  
 ATG CAG TGAAGCCCTC AGTGAGCTGC CACTGAGTCG GGAGCCCAGA GGAACGGCGT 3723  
 Met Gln  
 GAAGCCACTG GACCCCTCTCC CGGGACCTCC TGCTGGCTCC TCCTGCCCAG CTCCTATGC 3783  
 AAGGCTTTGT GTTTTCAGGA AAGGGCCTAG CTTCTGTGTG GCCCACAGAG TTCACTGCCT 3843  
 GTGAGGCTTA GCACCAAGTG CTGAGGCTGG AAGAAAAACG CACACCAGAC GGGCAACAAA 3903  
 CAGTCTGGGT CCCAGCTCG CTCCTGGTAC TTGGGACCCC AGTGCCTCGT TGAGGGCGCC 3963  
 ATTCTGAAGA AAGGAACTGC AGCGCCGATT TGAGGGTGGA GATATAGATA ATAATAATAT 4023  
 TAATAATAAT AATGGCCACA TGGATCGAAC ACTCATGATG TGCCAAGTGC TGTGCTAAGT 4083  
 GCTTTACGAA CATTCGTCAT ATCAGGATGA CCTCGAGAGC TGAGGCTCTA GCCACCTAAA 4143  
 ACACGTGCCC AAACCCACCA GTTTAAAACG GTGTGTGTTC GGAGGGGTGA AAGCATTAAG 4203  
 AAGCCCAGTG CCCTCCTGGA GTGAGACAAG GGCTCGGCCT TAAGGAGCTG AAGAGTCTGG 4263  
 GTAGCTTGTT TAGGGTACAA GAAGCCTGTT CTGTCCAGCT TCAGTGACAC AAGCTGCTTT 4323  
 AGCTAAAGTC CCGCGGGTTC CGGCATGGCT AGGCTGAGAG CAGGGATCTA CCTGGCTTCT 4383  
 CAGTTCTTTG GTTGGAAGGA GCAGGAAATC AGCTCCTATT CTCCAGTGGA GAGATCTGGC 4443  
 CTCAGCTTGG GCTAGAGATG CCAAGGCCTG TGCCAGGTTT CCTGTGCCCT CCTCGAGGTG 4503  
 GGCAGCCATC ACCAGCCACA GTTAAGCCAA GGGCCCCAAC ATGTATTCCA TCGTGCTGGT 4563  
 AGAAGAGTCT TTGCTGTTGC TCCCGAAAGC CGTGCTCTCC AGCCTGGCTG CCAGGGAGGG 4623  
 TGGGCCTCTT GGTTCAGGC TCTTGAAATA GTGCAGCCTT TTCTTCCTAT CTCTGTGGCT 4683  
 TTCAGCTCTG CTCCTTGGT TATTAGGAGA ATAGATGGGT GATGTCTTTC CTTATGTTGC 4743  
 TTTTCAACA TAGCAGAATT AATGTAGGGA GCTAAATCCA TGGGTGTGTG TGAATGCAGA 4803  
 AGGGAATGCA CCCACATTC CCATGATGGA AGTCTGCGTA ACCAATAAAT TGTGCCTTTC 4863  
 TTAAAAA 4870

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1187 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Val Pro Cys Trp Asn His Gly Asn Ile Thr Arg Ser Lys Ala Glu Glu  
 1 5 10 15  
 Leu Leu Cys Arg Thr Gly Lys Asp Gly Ser Phe Leu Val Arg Ala Ser  
 20 25 30

- 55 -

Glu Ser Ile Phe Arg Ala Tyr Ala Leu Cys Val Leu Tyr Arg Asn Cys  
                   35                                  40                                  45  
 Val Tyr Thr Tyr Arg Ile Leu Pro Asn Glu Asp Asp Lys Phe Thr Val  
                   50                                  55                                  60  
 Gln Ala Ser Glu Gly Val Ser Met Arg Phe Phe Thr Lys Leu Asp Gln  
   65                                  70                                  75                                  80  
 Leu Ile Glu Phe Tyr Lys Lys Glu Asn Met Gly Leu Val Thr His Leu  
                   85                                  90                                  95  
 Gln Tyr Pro Val Pro Leu Glu Glu Glu Asp Thr Gly Asp Asp Pro Glu  
                  100                                 105                                 110  
 Glu Asp Thr Glu Ser Val Val Ser Pro Pro Glu Leu Pro Pro Arg Asn  
                  115                                 120                                 125  
 Ile Pro Leu Thr Ala Ser Ser Cys Glu Ala Lys Glu Val Pro Phe Ser  
                  130                                 135                                 140  
 Asn Glu Asn Pro Arg Ala Thr Glu Thr Ser Arg Pro Ser Leu Ser Glu  
  145                                 150                                 155                                 160  
 Thr Leu Phe Gln Arg Leu Gln Ser Met Asp Thr Ser Gly Leu Pro Glu  
                  165                                 170                                 175  
 Glu His Leu Lys Ala Ile Gln Asp Tyr Leu Ser Thr Gln Leu Ala Gln  
                  180                                 185                                 190  
 Asp Ser Glu Phe Val Lys Thr Gly Ser Ser Ser Leu Pro His Leu Lys  
                  195                                 200                                 205  
 Lys Leu Thr Thr Leu Leu Cys Lys Glu Leu Tyr Gly Glu Val Ile Arg  
                  210                                 215                                 220  
 Thr Leu Pro Ser Leu Glu Ser Leu Gln Arg Leu Phe Asp Gln Gln Leu  
  225                                 230                                 235                                 240  
 Ser Pro Gly Leu Arg Pro Arg Pro Gln Val Pro Gly Glu Ala Asn Pro  
                  245                                 250                                 255  
 Ile Asn Met Val Ser Lys Leu Ser Gln Leu Thr Ser Leu Leu Ser Ser  
                  260                                 265                                 270  
 Ile Glu Asp Lys Val Lys Ala Leu Leu His Glu Gly Pro Glu Ser Pro  
                  275                                 280                                 285  
 His Arg Pro Ser Leu Ile Pro Pro Val Thr Phe Glu Val Lys Ala Glu  
  290                                 295                                 300  
 Ser Leu Gly Ile Pro Gln Lys Met Gln Leu Lys Val Asp Val Glu Ser  
  305                                 310                                 315                                 320  
 Gly Lys Leu Ile Ile Lys Lys Ser Lys Asp Gly Ser Glu Asp Lys Phe  
                  325                                 330                                 335  
 Tyr Ser His Lys Lys Ile Leu Gln Leu Ile Lys Ser Gln Lys Phe Leu  
                  340                                 345                                 350  
 Asn Lys Leu Val Ile Leu Val Glu Thr Glu Lys Glu Lys Ile Leu Arg  
                  355                                 360                                 365  
 Lys Glu Tyr Val Phe Ala Asp Ser Lys Lys Arg Glu Gly Phe Cys Gln  
                  370                                 375                                 380

SUBSTITUTE SHEET (RULE 26)

- 56 -

Leu 385	Leu	Gln	Gln	Met	Lys 390	Asn	Lys	His	Ser	Glu 395	Gln	Pro	Glu	Pro	Asp 400
Met	Ile	Thr	Ile	Phe 405	Ile	Gly	Thr	Trp	Asn 410	Met	Gly	Asn	Ala	Pro 415	Pro
Pro	Lys	Lys	Ile 420	Thr	Ser	Trp	Phe	Leu 425	Ser	Lys	Gly	Gln	Gly 430	Lys	Thr
Arg	Asp	Asp 435	Ser	Ala	Asp	Tyr	Ile 440	Pro	His	Asp	Ile	Tyr 445	Val	Ile	Gly
Thr	Gln	Glu	Asp	Pro	Leu	Ser	Glu 455	Lys	Glu	Trp	Leu 460	Glu	Ile	Leu	Lys
His 465	Ser	Leu	Gln	Glu	Ile 470	Thr	Ser	Val	Thr	Phe 475	Lys	Thr	Val	Ala	Ile 480
His	Thr	Leu	Trp	Asn 485	Ile	Arg	Ile	Val	Val 490	Leu	Ala	Lys	Pro	Glu	His 495
Glu	Asn	Arg	Ile 500	Ser	His	Ile	Cys	Thr 505	Asp	Asn	Val	Lys	Thr 510	Gly	Ile
Ala	Asn	Thr 515	Leu	Gly	Asn	Lys	Gly 520	Ala	Val	Gly	Val	Ser 525	Phe	Met	Phe
Asn 530	Gly	Thr	Ser	Leu	Gly	Phe 535	Val	Asn	Ser	His	Leu 540	Thr	Ser	Gly	Ser
Glu 545	Lys	Lys	Leu	Arg	Arg 550	Asn	Gln	Asn	Tyr	Met 555	Asn	Ile	Leu	Arg	Phe 560
Leu	Ala	Leu	Gly	Asp 565	Lys	Lys	Leu	Ser	Pro 570	Phe	Asn	Ile	Thr	His 575	Arg
Phe	Thr	His	Leu 580	Phe	Trp	Phe	Gly	Asp 585	Leu	Asn	Tyr	Arg	Val 590	Asp	Leu
Pro	Thr	Trp 595	Glu	Ala	Glu	Thr	Ile 600	Ile	Gln	Lys	Ile	Lys 605	Gln	Gln	Gln
Tyr 610	Ala	Asp	Leu	Leu	Ser	His 615	Asp	Gln	Leu	Leu	Thr 620	Glu	Arg	Arg	Glu
Gln 625	Lys	Val	Phe	Leu	His 630	Phe	Glu	Glu	Glu	Glu 635	Ile	Thr	Phe	Ala	Pro 640
Thr	Tyr	Arg	Phe	Glu 645	Arg	Leu	Thr	Arg	Asp 650	Lys	Tyr	Ala	Tyr	Thr 655	Lys
Gln	Lys	Ala	Thr 660	Gly	Met	Lys	Tyr	Asn 665	Leu	Pro	Ser	Trp	Cys 670	Asp	Arg
Val	Leu	Trp 675	Lys	Ser	Tyr	Pro	Leu 680	Val	His	Val	Val	Cys 685	Gln	Ser	Tyr
Gly 690	Ser	Thr	Ser	Asp	Ile	Met 695	Thr	Ser	Asp	His	Ser 700	Pro	Val	Phe	Ala
Thr 705	Phe	Glu	Ala	Gly	Val 710	Thr	Ser	Gln	Phe	Val 715	Ser	Lys	Asn	Gly	Pro 720
Gly	Thr	Val	Asp	Ser 725	Gln	Gly	Gln	Ile	Glu 730	Phe	Leu	Arg	Cys	Tyr 735	Ala

SUBSTITUTE SHEET (RULE 26)

Thr Leu Lys Thr Lys Ser Gln Thr Lys Phe Tyr Leu Glu Phe His Ser  
 740 745 750  
 Ser Cys Leu Glu Ser Phe Val Lys Ser Gln Glu Gly Glu Asn Glu Glu  
 755 760 765  
 Gly Ser Glu Gly Glu Leu Val Val Lys Phe Gly Glu Thr Leu Pro Lys  
 770 775 780  
 Leu Lys Pro Ile Ile Ser Asp Pro Glu Tyr Leu Leu Asp Gln His Ile  
 785 790 795 800  
 Leu Ile Ser Ile Lys Ser Ser Asp Ser Asp Glu Ser Tyr Gly Glu Gly  
 805 810 815  
 Cys Ile Ala Leu Arg Leu Glu Ala Thr Glu Thr Gln Leu Pro Ile Tyr  
 820 825 830  
 Thr Pro Leu Thr His His Gly Glu Leu Thr Gly His Phe Gln Gly Glu  
 835 840 845  
 Ile Lys Leu Gln Thr Ser Gln Gly Lys Thr Arg Glu Lys Leu Tyr Asp  
 850 855 860  
 Phe Val Lys Thr Glu Arg Asp Glu Ser Ser Gly Pro Lys Thr Leu Lys  
 865 870 875 880  
 Ser Leu Thr Ser His Asp Pro Met Lys Gln Trp Glu Val Thr Ser Arg  
 885 890 895  
 Ala Pro Pro Cys Ser Gly Ser Ser Ile Thr Glu Ile Ile Asn Pro Asn  
 900 905 910  
 Tyr Met Gly Val Gly Pro Phe Gly Pro Pro Met Pro Leu His Val Lys  
 915 920 925  
 Gln Thr Leu Ser Pro Asp Gln Gln Pro Thr Ala Trp Ser Tyr Asp Gln  
 930 935 940  
 Pro Pro Lys Asp Ser Pro Leu Gly Pro Cys Arg Gly Glu Ser Pro Pro  
 945 950 955 960  
 Thr Pro Pro Gly Gln Pro Pro Ile Ser Pro Lys Lys Phe Leu Pro Ser  
 965 970 975  
 Thr Ala Asn Arg Gly Leu Pro Pro Arg Thr Gln Glu Ser Arg Pro Ser  
 980 985 990  
 Asp Leu Gly Lys Asn Ala Gly Asp Thr Leu Pro Gln Glu Asp Leu Pro  
 995 1000 1005  
 Leu Thr Lys Pro Glu Met Phe Glu Asn Pro Leu Tyr Gly Ser Leu Ser  
 1010 1015 1020  
 Ser Phe Pro Lys Pro Ala Pro Arg Lys Asp Gln Glu Ser Pro Lys Met  
 1025 1030 1035 1040  
 Pro Arg Lys Glu Pro Pro Pro Cys Pro Glu Pro Gly Ile Leu Ser Pro  
 1045 1050 1055  
 Ser Ile Val Leu Thr Lys Ala Gln Glu Ala Asp Arg Gly Glu Gly Pro  
 1060 1065 1070  
 Gly Lys Gln Val Pro Ala Pro Arg Leu Arg Ser Phe Thr Cys Ser Ser  
 1075 1080 1085

- 58 -

Ser Ala Glu Gly Arg Ala Ala Gly Gly Asp Lys Ser Gln Gly Lys Pro  
 1090 1095 1100  
 Lys Thr Pro Val Ser Ser Gln Ala Pro Val Pro Ala Lys Arg Pro Ile  
 1105 1110 1115 1120  
 Lys Pro Ser Arg Ser Glu Ile Asn Gln Gln Thr Pro Pro Thr Pro Thr  
 1125 1130 1135  
 Pro Arg Pro Pro Leu Pro Val Lys Ser Pro Ala Val Leu His Leu Gln  
 1140 1145 1150  
 His Ser Lys Gly Arg Asp Tyr Arg Asp Asn Thr Glu Leu Pro His His  
 1155 1160 1165  
 Gly Lys His Arg Pro Glu Glu Gly Pro Pro Gly Pro Leu Gly Arg Thr  
 1170 1175 1180  
 Ala Met Gln  
 1185

**CLAIM:**

1. A purified and isolated nucleic acid molecule comprising a sequence encoding an SH2-containing inositol-phosphatase which has a src homology 2 (SH2) domain and exhibits phospholns-5-ptase activity.
- 5 2. An SH2-containing inositol-phosphatase as claimed in claim 1 which is further characterized by having an amino terminal src homology 2 (SH2) domain, two phosphotyrosine binding (PTB) consensus sequences, a proline rich region, and motifs highly conserved among inositol polyphosphate-5-phosphatases (phospholns-5-ptases).
3. A purified and isolated nucleic acid molecule as claimed in claim 1, comprising (i) a  
10 nucleic acid sequence encoding an SH2-containing inositol-phosphatase having the amino acid sequence as shown in SEQ ID NO:2 or Figure 2 (A); or, (ii) nucleic acid sequences complementary to (i).
4. A purified and isolated nucleic acid molecule as claimed in claim 1, comprising (i) a  
15 nucleic acid sequence encoding an SH2-containing inositol-phosphatase having the amino acid sequence as shown in SEQ ID NO:8 or Figure 11; or, (ii) nucleic acid sequences complementary to (i).
5. A purified and isolated nucleic acid molecule as claimed in claim 1, comprising (i) a  
nucleic acid sequence encoding an SH2-containing inositol-phosphatase having the nucleic acid  
sequence as shown in SEQ ID NO:1 or Figure 3, wherein T can also be U;  
20 (ii) a nucleic acid sequence complementary to (i); or  
(iii) a nucleic acid molecule differing from any of the nucleic acids of (i) and (ii) in  
codon sequences due to the degeneracy of the genetic code.
6. A purified and isolated nucleic acid molecule as claimed in claim 1, comprising (i) a  
nucleic acid sequence encoding an SH2-containing inositol-phosphatase having the nucleic acid  
25 sequence as shown in SEQ ID NO:7 or Figure 10, wherein T can also be U;  
(ii) a nucleic acid sequence complementary to (i); or  
(iii) a nucleic acid molecule differing from any of the nucleic acids of (i) and (ii) in  
codon sequences due to the degeneracy of the genetic code.
- 30 7. A purified and isolated nucleic acid molecule comprising a sequence which hybridizes  
under high stringency conditions to the nucleic acid molecule as claimed in claim 5.

8. A purified and isolated nucleic acid molecule as claimed in claim 1, which is a double stranded nucleic acid molecule or RNA.
9. A recombinant expression vector adapted for transformation of a host cell comprising a nucleic acid molecule as claimed in claim 1 and one or more transcription and translation  
5 elements operatively linked to the nucleic acid molecule.
10. A host cell containing a recombinant expression vector as claimed in claim 9.
11. A method for preparing an SH2-containing inositol-phosphatase comprising (a) transferring a recombinant expression vector as claimed in claim 9 into a host cell; (b) selecting transformed host cells from untransformed host cells; (c) culturing a selected transformed host  
10 cell under conditions which allow expression of the SH2-containing inositol-phosphatase; and (d) isolating the SH2-containing inositol-phosphatase.
12. A purified and isolated SH2-containing inositol-phosphatase which associates with Shc and exhibits phosphoIns-5-ptase activity.
13. A purified and isolated Shc protein as claimed in claim 12, which has the amino acid  
15 sequence as shown in SEQ ID NO:2 or Figure 2(A), or as shown in SEQ ID NO:8 or Figure 11.
14. Antibodies having specificity against an epitope of the SH2-containing inositol-phosphatase as claimed in claim 13.
15. A nucleotide probe comprising a sequence encoding at least 6 continuous amino acids from the SH2-containing inositol-phosphatase as shown in SEQ ID. NO. 2 or Figure 2(A), or  
20 as shown in SEQ ID. NO. 8 or Figure 11.
16. A method for identifying a substance which is capable of binding to a purified and isolated SH2-containing inositol-phosphatase protein as claimed in claim 12, comprising reacting the protein with at least one substance which potentially can bind with the protein under conditions which permit the formation of complexes between the substance and the  
25 protein; and, assaying for complexes, for free substance, for non-complexed protein, or for activation of the protein.
17. A method for assaying a medium for the presence of an agonist or antagonist of the interaction of a purified and isolated SH2-containing inositol-phosphatase protein as claimed  
30 in claim 12 and a substance which binds to the protein which comprises reacting the protein

with a substance which is capable of binding to the protein and a suspected agonist or antagonist substance, under conditions which permit the formation of complexes between the substance and the protein; and, assaying for complexes, for free substance, for non-complexed protein, or for activation of the protein.

5 18. A method as claimed in claim 17, wherein the substance is Shc or a part thereof.

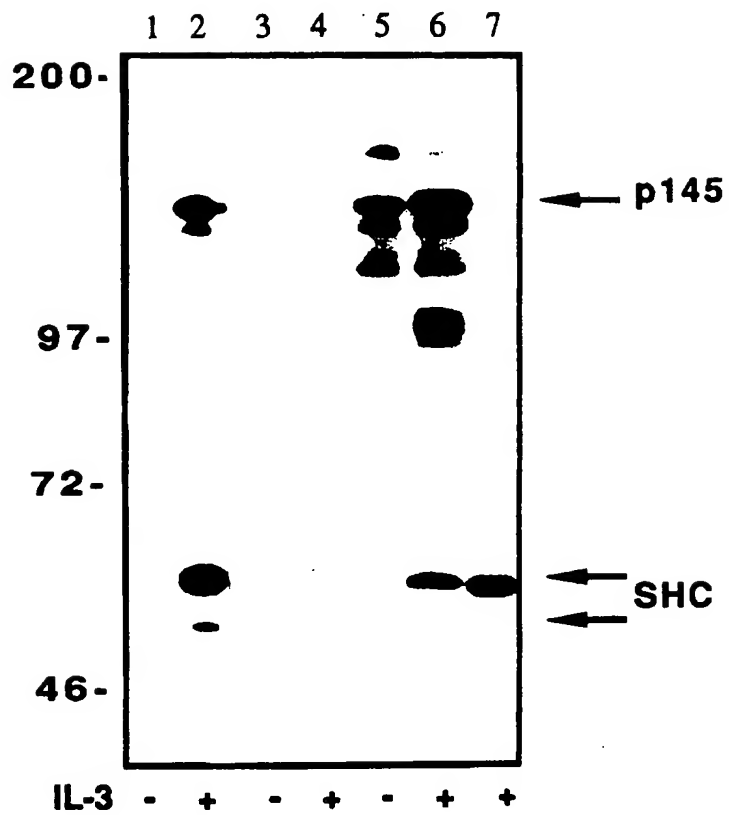
19. A method for assaying for the affect of a substance on the phosphoIns-5-ptase activity of a SH2-containing inositol-phosphatase protein as claimed in claim 12 comprising reacting a substrate which is capable of being hydrolyzed by the protein to produce a hydrolysis product,  
10 with a substance which is suspected of affecting the phosphoIns-5-ptase activity of the protein, under conditions which permit the hydrolysis of the substrate; determining the amount of hydrolysis product; and, comparing the amount of product obtained with the amount obtained in the absence of the substance to determine the affect of the substance on the phosphoIns-5-ptase activity of the protein.

15

20. A substance identified in accordance with the method of claim 16, 17, 18 or 19.

21. A pharmaceutical composition comprising a substance identified in accordance with the method of claim 16.

1/27  
**FIGURE 1**



2/27  
**FIGURE 2**

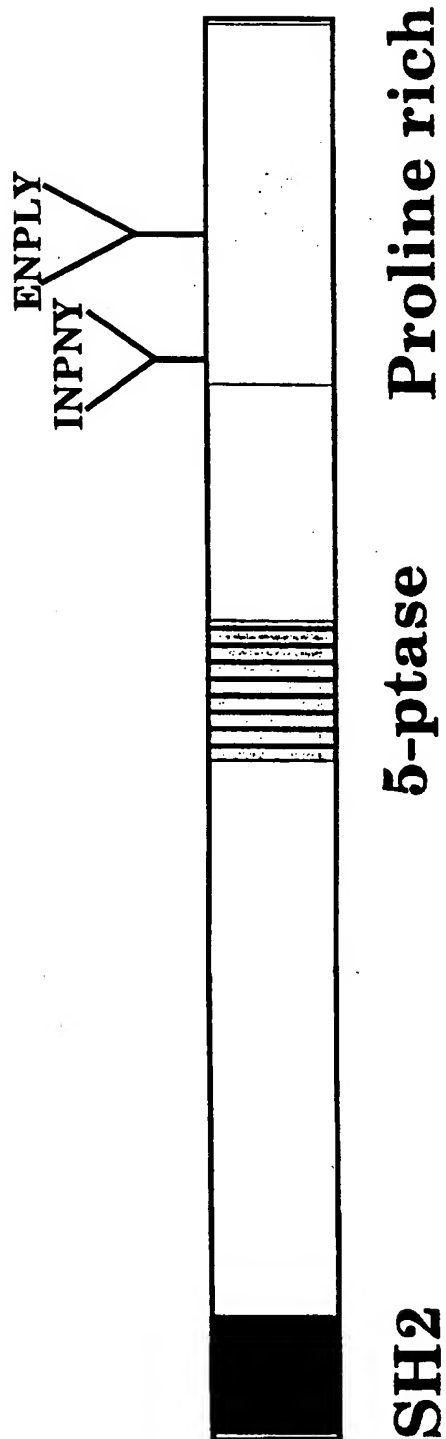
**A**

1 MPAMVPG WNHGNITRSKAEELLSRAGKDGSLVRASESIPRACALCVLFR  
 51 NCVYTYRILPNEDDKFTVQASEGVPMRFFTKLDQLIDFYKKENMGLVTHL  
 101 QYPVP LEEEDAIDEAEEDTESVMSPPELPPRNIPMSAGPSEAKDLPLATE  
 151 NPRAPEVTRLSLSETLFQRLQSMDSGLPEEHLKAIQDYLSTQLLDSDF  
 201 LKTGSSNLPHLKKLMSLLCKELHGEVIRTLPSLESLQRLFDQQLSPGLRP  
 251 RPQVPGEASPITMVAKLSQLTSLLSSIEDKVKSLLHEGSESTNRRSLIPP  
 301 VTFEVKSESLGIPQKMHLKVDVESGKLIVKSKDGSSEDKFYSHKKILQLI  
 351 KSQKFLNKLVLVETEKEKILRKEYVFADSKKREGFCQLLQQMKNKHSEQ  
 401 PEPDMITIFIGTWNMGNAPPPKKITSWFLSKGQGKTRDDSADYIPHDIVV  
 451 IGTQEDPLGEKEWLELLRHSLQEVTSMFTKTVAIHTLWNIRIVVLAKPEH  
 501 ENRISHICTDNVKTGIANTLGNGKAVGVVSFMFNGTSLGFVNSHLTSGSEK  
 551 KLRRNQNYMNLRFALGDKKLSPFNTHRFTHLFWLGDLNRYVELPTWE  
 601 AEAIIQKIKQQQYSDLLAHDQLLLERKDQKVFLHFEEEEITFAPTYRFER  
 651 LTRDKYAYTKQKATGMKYNLPSWCDRVLWKSYPVHVVCQSYGSTSDIMT  
 701 SDHSPVFATFEAGVTSQFVSKNGPGTVDSQGQIEFLACYATLTKTSQTKF  
 751 YLEFHSSCLESFVKSQEGENESEGEVVRFGETLPKLKPIISDPEYLL  
 801 DQHILISIKSSDSDESYGEGCIALRLETTEAQHPITYTPLTHHGEMTGHR  
 851 GEIKLQTSQGKMREKLYDFVKTERDESSGMKCLKNLTSHDPMRQWEPGR  
 901 VPACGVSSLNEMINPNYIGMGPFQPLHGKSTLSPDQQLTAWSYDQLPKD  
 951 SSLGPGRGEGPPTPPSPPLSPKKFSSSTNRRGPCPRVQEARPGDLGKVE  
 1001 ALLQEDLLLTKPEMFENPLYGVSSSF PKL VPRKEQESPKMLRKEPPPCPD  
 1051 PGISSPSIVLPKAEVESVKGTSKQAPVPVLGPTPRIRSFTCSSSAEGRM  
 1101 TSGDKSQGKPKASASSQAPVPVKRPVKPSRSEMSQQTTPAPRPPLPVK  
 1151 SPAVLQLQHSKGRDYRDNTELP HHGKHRQEEGLLGRTAMQ

3/27

FIGURE 2 CONT'D

B



4/27

**FIGURE 3**

&gt;BASE COUNT 1014 a 1147 c 1054 g 825 t

&gt;ORIGIN

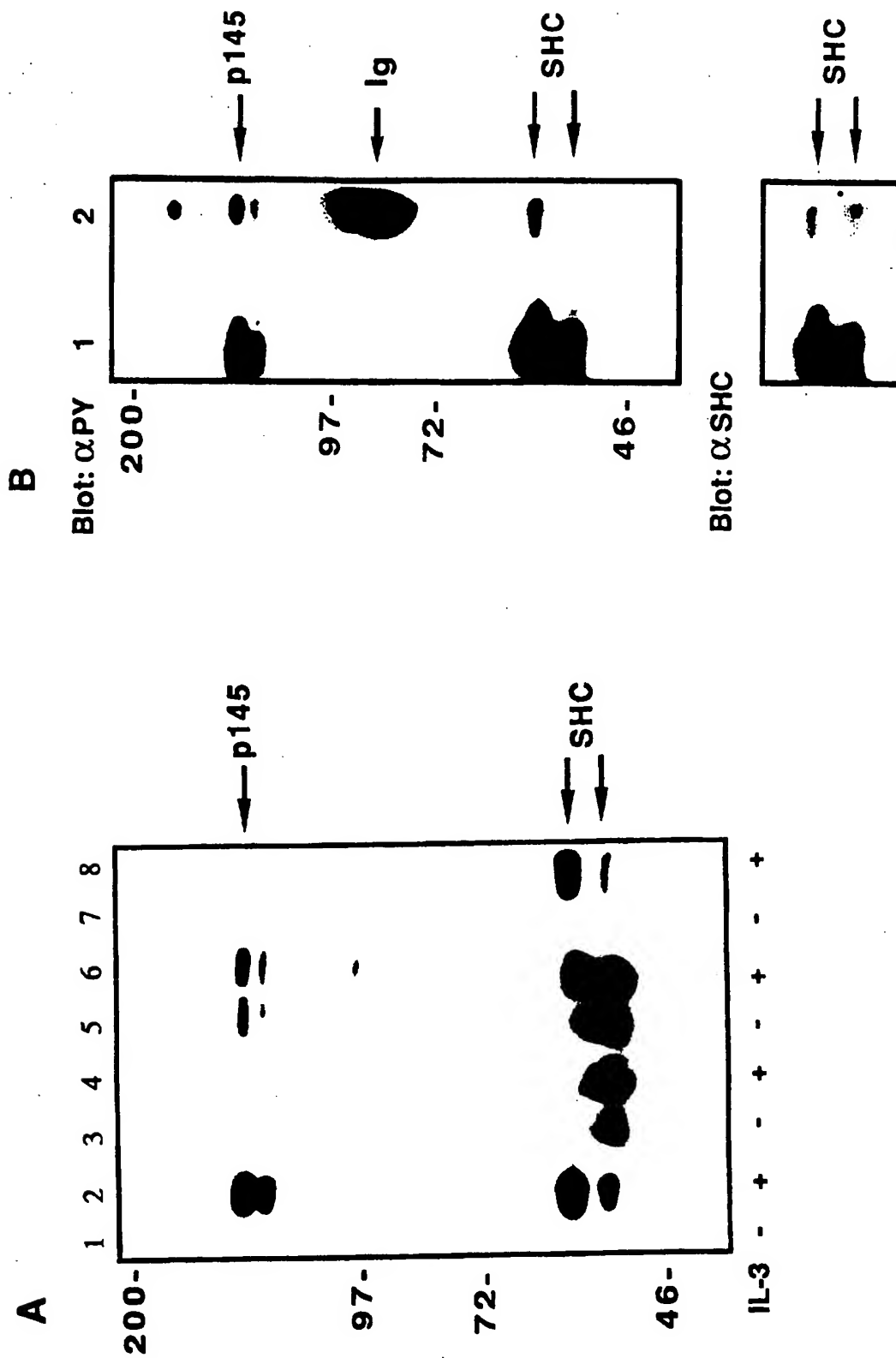
```

> 1 ccttgtagg agcagcagag gcaattctg agaggcaaca ggccggcagg ctcagcctag
> 61 agagggccct gaactacttt gctggaggt ccttctggg agtggctgct gacccagtc
> 121 aggagaccca tgcctgccat ggtccctggg tggaaacatg gcaacatcac ccgctccaag
> 181 gcagaggagc tactttccag agccggcaag gacgggagct tcttctgctg tgcagcagag
> 241 tccatccccc ggccctgcgc actctgcgtg ctgtccgga atgtgttta cacttacagg
> 301 attctgcca atgaggacga taaattcact gticaggcat ccgaagggtt ccccatgagg
> 361 ttctcacga agctggacca gctcatgac tttaacaaga aggaaaacat ggggctggg
> 421 acccacttgc agtaccocgt gccctggag gaggaggatg ctattgatga ggctgaggag
> 481 gacactgaaa gtgtcatgtc accacctgag ctgcctcca gaaacatcc tatgtctgc
> 541 gggccacgag agggcaagga ccttctctt gcaacagaga acccccgagc ccttgaggtc
> 601 acccggtcga gtctctccga gacactgttt cagcgtctac agagcatgga taccagtggg
> 661 cttcccgagg agcacctgaa agccatccag gattatctga gcactcagct cctctggat
> 721 tccgactttt tgaaaacggg ctccagcaac ctccctcacc tgaagaaagt gatgtcactg
> 781 ctctgcaagg agctccatgg ggaagtcac aggaacttgc catccctgga gtctctcag
> 841 aggtgtttg accaacagct ctcccaggc ctctgcccac gaactcaggt gcccgagag
> 901 gccagtccca tcaccatggt tgcacaactc agccaatga caagtctgct gtcttccat
> 961 gaagataagg tcaagtcctt gctgcacgag ggctcagaat ctaccaacag gcgttccct
> 1021 atccctccgg tcacctttga ggtgaagta ggttccctgg gcatctcca gaaaatgcat
> 1081 ctcaaagtgg acgttgagtc tgggaaactg atcgttaaga agtccaagga tgggtctgag
> 1141 gacaagtctc acagccacaa aaaaatcctg cagctcatta agtccagaa gtttctaac
> 1201 aagtgtgtga ttgtgtgga gacggagaag gaaaaatcc tgaggaaagga atagtgttt
> 1261 gctgactcta agaaaagaga aggtctctg caactctgc agcagatgaa gaacaagcat
> 1321 tgggagcagc cagagcctga catgatcacc atctcatg gcacttgaa calgggtaat
> 1381 gcacccctc ccaagaagat cagctctgg ttctctcca aggggcaggg aaagacacgg
> 1441 gacgactctg ctgactacat ccccatgac atctatgta ttggcacca ggaggatcc
> 1501 ctggagaga aggagtggct ggagctactc aggcactccc tgaagaagt caccagcatg
> 1561 acatttaaaa cagttgccat ccacaccctc tgaacattc gcatagtgtt gcttgccaag
> 1621 ccagagcatg agaatcggat cagccatac tgcactgaca acgtgaagac aggcacgccc
> 1681 aacacccctg gaaacaaggg agcagtgga gtgtcttca tgtcaatgg aacctcttg
> 1741 gggttctga acagccactt gactctgga agtgaaaaa agtcaggag aaatcaaac
> 1801 tatatgaaca tctgcgggtt cctggccctg ggagacaaga agctaagccc atttaacatc
> 1861 acccaccgct tcacccacct ctcttggtt ggggatctca actaccgct ggagctgccc
> 1921 acttgaggag cagaggccat catccagaag atcaagcaac agcagtatc agacctctg
> 1981 gccacgacc aactgctct ggagaggaa gaccagaagg tcttctgca cttgaggag
> 2041 gaagagatca ccttgcgcc cactatcga ttgaagac tgaccggga caagtatga
> 2101 tacagaaac agaaagcaac agggatgaag taactatgc cgtctggtg cgaccgagtc
> 2161 ctctggaagt ctaccgctt ggtgcatgtg gtctgtcagt cctatggcag taccagtac
> 2221 atcatgacga gtgaccacag cctgtctt gccagttt agcaggagt cacatctca
> 2281 ttctctcca agaattgtcc tggcactgta gatagccaag ggcagatcga gtttctgca
> 2341 tgcctagcca cactgaagac caagtcacc actaagtct acttgaggt ccaactcaagc
> 2401 tgcctagaga gtttctca ggtcaggaa ggagagaatg aagagggaag tgaaggagag
> 2461 ctgtgtgtac ggttgagga gactctcc aagctaaagc ccatatctc tgaccocgag
> 2521 tactactgg accagcatat cctgacagc attaaatct ctgacgtga cagtgctat
> 2581 ggtgaagggt gcatgccc tgcctggag accacagagg ctacagatcc tatctacacg
> 2641 cctctacccc accatgggga gatgactgg cacttcagg gagagattaa gctgcagacc
> 2701 tccagggca agatgaggga gaagcttat gactttga agacagagcg ggaatgaatc
> 2761 agtgaatga aatgctgaa gaacctcacc agccatgacc ctatgaggca atgggagcct
> 2821 tctggcaggg tccctgcatg tgggtctcc agctcaatg agatgatca tccaaactac
> 2881 attggtatg ggccttttg acagccctg catgggaaat caacctgtc ccagatcag
> 2941 caactcacag ctggaggtta tgaccagta ccaaaagact cctccctggg gcttgggagg
> 3001 ggggagggtc ctcaacccc tccctccaa ccaactctg cgcgaagaa gtttcatct
> 3061 tcacaccca accgaggtcc ctgcccagg gtgcaagg caagacctg ggaatctgga
> 3121 aaggtggaag ctctgtcca ggaggacct ctgtgacga agcccgagat gttgagaac
> 3181 ccactgtatg gatccgtgag ttcttccct aagctgtgtc ccaggaaaga gcaggagtct
> 3241 ccaagatgc tgcggaagga gcccggccc tgtccagacc caggaaatc atcacccagc
> 3301 atcgtgtccc ccaaaagccc agaggtggag agtgtcaagg ggacaagcaa acaggccctt
> 3361 gtgctgtccc ttggcccac accccggatc cgtcttita cctgttctc ttctgtgag
> 3421 ggcagaatga ccagtggga caagagcca ggaagcca aggcctcagc cagtcccaa
> 3481 gcccagtcg cagtacagag gctgtcaag ccttcaggt cagaaatgag ccagcagaca
> 3541 acacccatcc cagctccagc gccacccctg ccagtcaaga gtctgtctg cctgcagctg
> 3601 caacatcca aagcgagaga ctaccgtgac aacacagaac tccccacca tggcaagcac
> 3661 cgccaagagg aggggctgct tggcaggact gccatgagt gactgtctg tgaatggagc
> 3721 ctggaggaa agcacaagc agacctgga cctcttcag gatgctctc tcaagatgcc
> 3781 tctggagga octctgtcta gctctctg ctagcttca agtccaggc tgtgtattt
> 3841 tttaaggaa accgctcac ttctgtgg tcaagaagt gtgtgtctg ctgccacat
> 3901 gtgcggcaga tgcataagct ggaatcaaaa cgcagccat acagacagca gacagcggca
> 3961 ctgggttca gaacttggat tcttgggct tcttcagtc gccgtttta agaaaggaa
> 4021

```

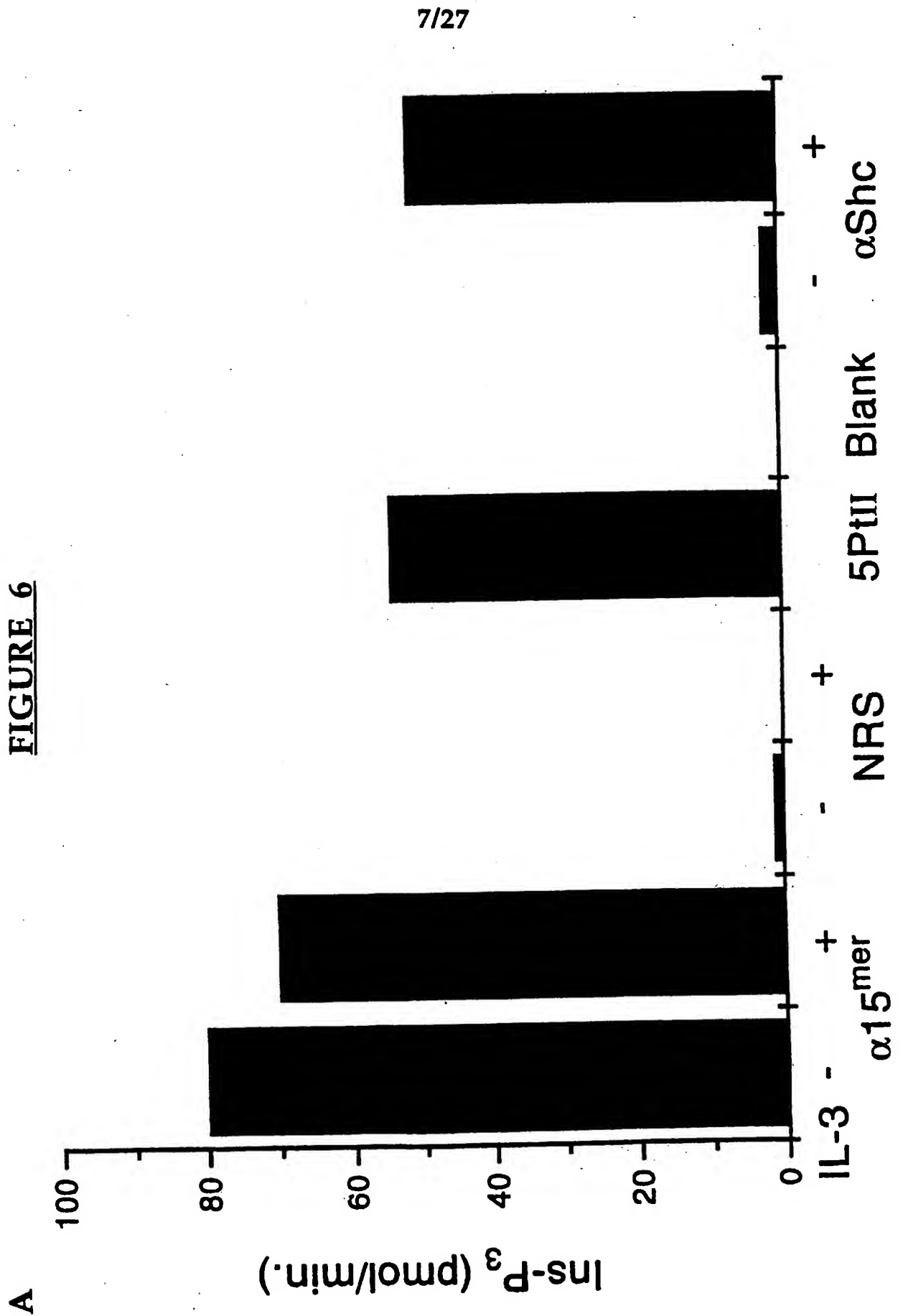
5/27

**FIGURE 4**

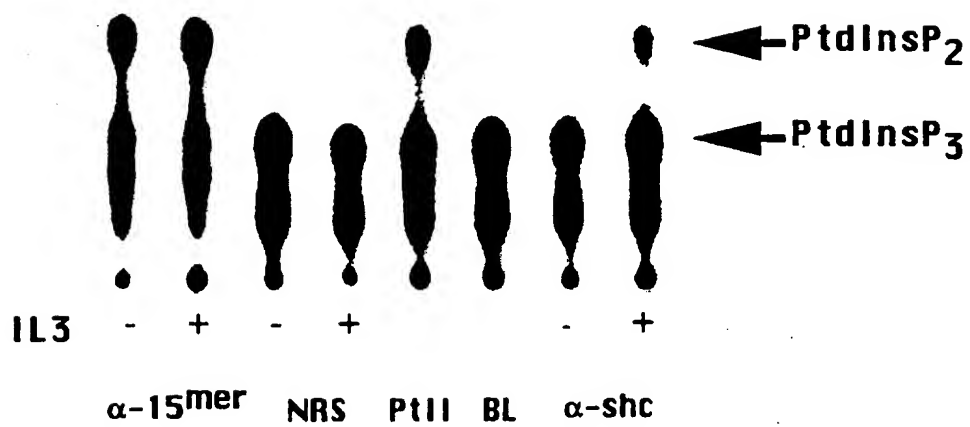


6/27  
**FIGURE 5**



**FIGURE 6**

8/27  
**FIGURE 6 CONT'D**

**B**

9/27  
**FIGURE 7**

Gene	Locus: SHC1	gil134475: 1..473
Organism	HOMO SAPIENS (HUMAN)	gil134475: 1..473
Sequence	473 aa	

```
1 mnklsggggr rtrveggqlg geewtrhgsf vnkptrgwlh pndkvmgpgv
51 sylvrymgcv evlqsmrald fntrtqvtr eaislvceavp gakgatrirk
101 pcsrplssil grsnlkfagm pitltvstss lnlnmaadckq iianhhmqsi
151 sfasggdpdt aeyvayvakd pvnqrachil ecpeglaqdv istigqafel
201 rfkqylrnpp klvtphdrma gfdgsawdee eeppdhqyy ndfpgkeppl

251 ggvdmlre gaapgaarpt apnaqtpshl gatlpvgqp ggdpvirkqm
301 pppppcpgre lfddpsvvnv qnldkarqav ggagppnpai ngsaprdlfd
351 mkpfedalrv ppppqsvsma eqlrgepwfh gklrreaea llqlngdflv
401 resttppgqy vltglqsgqp khllldpeg vvrtdhrfe svshlisym
451 dnhlpiisag selclqqpve rkl
```

10/27  
**FIGURE 8**

H.sapiens SHC mRNA.

ACCESSION X68148

\*FIELD\* NID

g36453

KEYWORDS SHC protein.

SOURCE human.

ORGANISM Homo sapiens

Eukaryotae; mitochondrial eukaryotes; Metazoa/Eumycota group;  
Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata;  
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;  
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;  
Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 3031)

AUTHORS Pelicci,P.

TITLE Direct Submission

JOURNAL Submitted (10-JUN-1992) to the EMBL/GenBank/DDBJ databases. P.  
Pelicci, Clinica Medica I, Policlinico Montelucente, Perugia 06100  
08854, ITALY

REFERENCE 2 (bases 1 to 3031)

AUTHORS Pelicci,G., Lanfranccone,L., Grignani,F., McGlade,J., Cavallo,F.,  
Forni,G., Nicoletti,I., Grignani,F., Pawson,T. and Pelicci,P.G.

TITLE A novel transforming protein (SHC) with an SH2 domain is implicated  
in mitogenic signal transduction

JOURNAL Cell 70 (1), 93-104 (1992)

MEDLINE 92323554

FEATURES Location/Qualifiers

source 1..3031  
/organism="Homo sapiens"

CDS 82..1503  
/codon\_start=1  
/product="SHC transforming protein"  
/db\_xref="PID:g36454"

/translation="MNKLSGGGGRTRVEGGQLGGEEWTRHGSFVNKPTRGW  
LHPNDK

VMGPGVSYLVRYMGCVEVLQSMRALDFNTRTQVTREASLVCEAVPGAKGATR  
RRKPC

SRPLSSILGRSNLKFAGMPITLTVSTSSLNLMAADCKQIIANHHMQSISFASGGDPD  
T

AEYVAYVAKDPVNQRACHILECPEGLAQDVISTIGQAFELRFKQYLRNPPKLVTPH  
DR

MAGFDGSAWDEEEEEPPDHQYYNDFPGKEPPLGGVDMRLREGAAPGAARPTAP

11/27

**FIGURE 8 CONT'D**

NAQT

PSHLGATLPVGQPVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKARQAV  
GGAG

PPNPAINGSAPRDLFDMKPFEDALRVPPPPQSVSMAEQLRGEPWFH GKLSRREAE  
ALL

QLNGDFLVRESTTTPGQYVLTGLQSGQPKHLLLVDPEGVVRTKDH RFESVSHLISY  
HM

DNHLPISAGSELCLQQPVERKL"

BASE COUNT 664 a 855 c 809 g 703 t

ORIGIN

1 gcggtaacct aagctggcag tggcgtgac cggcaccaaa tcggcccgcg gtgcgtgcgg  
61 agactccatg aggccctgga catgaacaag ctgagtggag gcggcgggag caggactcgg  
121 gtggaaggagg gccagcttgg gggcgaggag tggaccgcc acgggagctt tgtcaataag  
181 cccacgcggg gctggctgca tccaacgac aaagtcattg gacccggggt ttctacttg  
241 gttcggtaca tgggttgtgt ggaggtctc cagtcaatgc gtgccctgga ctcaacacc  
301 cggactcagg tcaccaggga ggccatcagt ctggtgtgtg aggtctgcc ggggtgtaag  
361 ggggagcaaa ggaggagaaa gccctgtagc cggccgtca gctctatctt ggggaggagt  
421 aacctgaaat ttgctggaat gccaatcact ctaccgtct ccaccagcag cctcaacctc  
481 atggccgcag actgcaaaca gatcatcgcc aaccaccaca tgcaatctat ctcatattga  
541 tccggcgggg atccggacac agccgagtat gtcgcctatg ttgcaaaga ccctgtgaat  
601 cagagagcct gccacattct ggagtgtccc gaagggttg ccaggatgt catcagcacc  
661 attggccagg ctttcgagtt gcgcttcaa caataacct ggaaccacc caaactggtc  
721 accctcatg acaggatggc tggctttgat ggctcagcat gggatgagga ggaggaagag  
781 ccactgacc atcagtacta taatgactt cgggggaagg aacccccctt ggggggggtg  
841 gtagacatga ggcttcggga aggagccgt ccaggggctg ctgacccac tgcaccaat  
901 gccagaccc ccagccactt gggagctaca ttgctgtag gacagcctgt tgggggagat  
961 ccagaagtcc gcaaacagat gccacctca ccacctgtc caggcagaga gcttttgat  
1021 gatccctct atgtcaact ccagaacct gacaaggccc ggcaagcagt ggggtgtgct  
1081 gggccccca atctgctat caatggcagt gacccccggg acctgttga catgaagccc  
1141 ttcgaagatg ctctcgggt gcctccacct cccagtcgg tgtccatggc tgagcagtc  
1201 cgaggggagc cctggttcca tgggaagctg agccggcggg aggtgaggc actgtgcag  
1261 ctcaatgggg acttcttgt acgggagagc acgaccacac ctggccagta tgtgtcact  
1321 ggcttgca ga tggggcagcc taagcatttg ctactggttg acctgaggg tgtggttcgg  
1381 actaaggatc accgtttga aagtgtcagt cacctatca gctaccacat ggacaatcac  
1441 ttgcccata tctctcggg cagcgaactg tgtctacag aacctgtgga gcggaaactg  
1501 tgatctgccc tagcgtctc ttccagaaga tgccctcaa tctttccac cctattccct  
1561 aactctcggg acctggttg ggagtgtct gtgggcttgg cttgtgtca gagctgggag  
1621 tagcatggac tctgggttc atatccagct gagtgagagg gtttgagta aaagcctggg  
1681 tgagaatcct gcctctccc aaacattaat caccaaagta ttaatgtaca gagtggcccc  
1741 tcactgggc cttctgtg ccaacctgat gcccttccc caagaagggt agtgcttgc  
1801 atggaaaatg tctgtggtg acaggccag tgaacagtc accttctgg gcaaggggga  
1861 acaaatcaca cctctgggt tcagggtatc ccagaccct ctcaacacc gccccccca  
1921 tgtttaaact ttgtgcctt gaccatctct taggtcta at gatatttat gcaaacagt

12/27

**FIGURE 8 CONT'D**

1981 cttggacccc tgaattcttc aatgacaggg atgccaacac cttcttggct tctgggacct  
2041 gtgttcttgc tgagcacct ctccggttg ggttgggata acagaggcag gagtggcagc  
2101 tgtccctct cctggggat atgcaacct tagagattgc ccagagccc cactcccggc  
2161 caggcgggag atggaccct ccttgcctca gtgcctctg gccggggccc ctcacccaa  
2221 ggggtctgta tatacttcc ataaggcctg cctcccatg ttgcatgcct atgtactctg  
2281 cgccaaagt cagccctcc tctgaagcc tctgcctgc ctcccttct gggagggcgg  
2341 ggtgggggtg actgaattg ggcctctgt acagttaact ctcccaggtg gattttgtg  
2401 aggtgagaaa aggggcattg agactataaa gcagtagaca atccccacat accatctgta  
2461 gaggttgaac tgcattctt taaagttta tatgcatata ttttagggct gctagactta  
2521 ctttctatt ttctttcca ttgcttatt ttgagcaca aatgataatc aattattaca  
2581 ttatacatc accttttga ctttccaag ccttttaca gctcttggca ttttctcgc  
2641 ctaggcctgt gaggttaactg ggatcgacc ttltatacca gagacctgag gcagatgaaa  
2701 ttattttcca tctaggacta gaaaaactg ggtctcttac cgcgagactg agaggcagaa  
2761 gtcagcccg atgcctgtca gtttcatgga ggggaaacgc aaaacctgca gttctgagt  
2821 accttctaca ggcccggccc agcctaggcc cggggtggcc acaccacagc aagccggccc  
2881 cccctcttt ggccttgtg ataagggaga gttgaccgtt tcatcctgg cctcctttg  
2941 ctgtttggat gttccacgg gtctcactta taccaaaggg aaaactcttc attaaagtcc  
3001 cgtatttctt ctaaaaaaaa aaaaaaaaaa a

//

13/27  
**FIGURE 9**

NCBI gi: 181975

FEATURES                      Location/Qualifiers

    source                      1..1109

                                /organism="Homo sapiens"

                                /sequenced\_mol="cDNA to mRNA"

                                /tissue\_type="brainstem"

                                /tissue\_lib="gt11 human brainstem library"

    CDS                         79..732

                                /gene="EGFRBP-GRB2"

                                /note="NCBI gi: 181976"

                                /codon\_start=1

                                /product="epidermal growth factor receptor-binding protein GRB2"

                                /translation="MEAIKYDFKATADDELSFKRGDILKVLNEECDQNWYKAELNGK  
DGFIPKNYIEMKPHPWFFGKIIPRAKAEEMLSKQRHDGAFLIRESESAPGDFSLSVKFG  
NDVQHFKVLRDAGAGKYFLWVVKFNSLNELVDYHRSTSVSRNQIFLRDIEQVPQOPTY  
VQALFDFDPQEDGELGFRRGDFIHVMDNSDPNWWKGACHGQTGMFPRNYVTPVNRNV"

BASE COUNT                  313 a       273 c       262 g       261 t

ORIGIN

```

1 gccagtgaat tcggggggctc agccctctctc cctcccttcc ccttgcttca ggctgctgag
61 cactgagcag cgctcagaat ggaagccatc gccaaatatg acttcaaagc tactgcagac
121 gacgagctga gcttcaaaag gggggacatc ctcaagggtt tgaacgaaga atgtgatcag
181 aactggtaca aggcagagct taatggaaaa gacggcttca ttcccaagaa ctacatagaa
241 atgaaaccac atccgtggtt ttttggcaaa atccccagag ccaaggcaga agaaatgctt
301 agcaaacagc ggcacgatgg ggccttttctt atccgagaga gtgagagcgc tcctggggac
361 ttctccctct ctgtcaagtt tggaaacgat gtgcagcact tcaagggtgt ccgagatgga
421 gccgggaagt acttcctctg ggtggtgaag ttcaattctt tgaatgagct ggtggattat
481 cacagatcta catctgtctc cagaaaccag cagatattcc tgcgggacat agaacagggtg
541 ccacagcagc cgacatacgt ccaggccctc tttgactttg atccccagga ggatggagag
601 ctgggcttcc gccggggaga ttttatccat gtcatggata actcagaccc caactggtgg
661 aaaggagctt gccacgggca gaccggcatg tttccccgca attatgtcac ccccgatgaac
721 cggaacgtct aagagtcaag aagcaattat ttaaagaaag tgaaaaatgt aaaacacata
781 caaaagaatt aaaccacaaa gctgcctctg acagcagcct gtgagggagt gcagaacacc
841 tggccgggtc accctgtgac cctctcactt tggttggaac tttagggggt gggagggggc
901 gttggattta aaaatgccaa aacttaccta taaattaaga agagttttta ttacaaattt
961 tcaactgctg tcctctttcc cctcctttgt cttttttttc atcctttttt ctcttctgtc
1021 catcagtgca tgacgtttta ggccacgtat agtcctagct gacgccaata ataaaaaaca
1081 agaaacaaaa aaaaaaaaaa ccgaattca

```

//

14/27  
**FIGURE 10**

**hSHIP cDNA Sequence**

5' UNTRANSLATED REGION (1-128)

1	GAATTCGCGG	CCGCTCGAC	CCAAGAGGCA	ACGGGCGGCA	GGTTGCAGTG	
51	GAGGGGCTC	CGCTCCCCC	GGTGGTGTGT	GGGTCCCTGGG	GGTGGCTGCC	
101	GGCCCAGCCG	AGGAGGCCCA	CGCCCAGCAT	GGTCCCCTGC	TGGAACCATG	START CODON
151	GCAACATCAC	CCGCTCCAAG	GCGGAGGAGC	TGCTTTGCAG	GACAGGCAAG	
201	GACGGGAGCT	TCCTCGTGCG	TGCCAGCGAG	TCCATCTTCC	GGGCATACGC	
251	GCTCTGCGTG	CTGTATCGGA	ATTGCGTTTA	TACTTACAGA	ATTCTGCCCA	
301	ATGAAGATGA	TAAATTCAT	GTTTCAGGCAT	CCGAAGGCGT	CTCCATGAGG	
351	TTCTTCACCA	AGCTGGACCA	GCTCATCGAG	TTTTACAAGA	AGGAAAACAT	
401	GGGGCTGGTG	ACCCATCTGC	AATACCCTGT	GCCGCTGGAG	GAAGAGGACA	
451	CAGCGGACGA	CCCTGAGGAG	GACACAGAAA	GTGTCGTGTC	TCCACCCGAG	
501	CTGCCCCCAA	GAAACATCCC	GCTGACTGCC	AGCTCCTGTG	AGGCCAAGGA	
551	GGTTCTTTT	TCAAACGAGA	ATCCCCGAGC	GACCGAGACC	AGCCGGCCGA	
601	CCCTCTCCGA	GACATTGTTT	CAGCGACTGC	AAAGCATGGA	CACCAAGTGG	
651	GCTTCCAGAAG	AGCATCTTAA	GGCCATCCAA	GATTATTTAA	GCACTCAGCT	
701	CGCCCAGGAC	TCTGAATTTG	TGAAGACAGG	GTCCAGCAGT	CTTCTCACC	
751	TGAAGAACT	GACCACACTG	CTCTGCAAGG	AGCTCTATGG	AGAAGTCATC	
801	CGGACCCTCC	CATCCCTGGA	GTCTCTGCAG	AGGTTATTTG	ACCAGCAGCT	
851	CTCCCCGGG	CTCCGTCCAC	GTCTCTCAGT	TCCGTGGTGA	GCCAATCCCA	
901	TCAACATGGT	GTCCAAGCTC	AGCCAACCTA	CAAGCCTGTT	GTCTATCCATT	
951	GAAGACAAGG	TCAAGGCCTT	GCTGCACGAG	GGTCTGAGT	CTCCGCACCG	
1001	GCCCTCCCTT	ATCCCTCCAG	TCACCTTTGA	GGTGAAGGCA	GAGTCTCTGG	
1051	GGATTCTCTA	GAAATGTCAG	CTCAAGTCTG	ACGTTGAGTC	TGGGAAACTG	
1101	ATCATTAAAG	AGTCCAAGGA	TGGTTCTGAG	GACAAGTTCT	ACAGCCACAA	
1151	GAAATTCCTG	CAGCTCATTA	AGTCACAGAA	ATTCTGGAAT	AAGTTGGTGA	
1201	TCTTGGTGGG	AACAGAGAAG	GAGAAGATCC	TCCGGAAGGA	ATATGTTTTT	
1251	GCTGACTCCA	AAAAGAGAGA	AGGCTTCTGC	CAGCTCCTGC	AGCAGATGAA	
1301	GAACAAGCAC	TCAGAGCAGC	CGGAGCCCGA	CATGATCACC	ATCTTCATCG	
1351	GCACCTGGAA	CATGGGTAAC	GCCCCCCTC	CCAAGAAGAT	CACGTCTCTG	
1401	TTTCTCTCCA	AGGGGCAGGG	AAAGACGCGG	GACGACTCTG	CGGACTACAT	
1451	CCCCCATGAC	ATTTACGTGA	TGGGCACCCA	AGAGGACCCC	CTGAGTGAGA	
1501	AGGAGTGGCT	GGAGATCCTC	AAACACTCCC	TGCAAGAAAT	CACCACTCTG	
1551	ACTTTTAAAA	CAGTCGCCAT	CCACACGCTC	TGGAACATCC	GCATCGTGGT	
1601	GCTGGCCAAAG	CCTGAGCAGG	AGAACCAGAT	CAGCCACATC	TGTAAGTACA	
1651	ACGTGAAGAC	AGGCATTGCA	AACACACTGG	GGAACAAGGG	AGCCGTGGGG	
1701	GTGTCGTTCA	TGTTCAATGG	AACCTCCTTA	GGGTTCTGTA	ACAGCCACTT	
1751	GACTTCAGGA	AGTGAAAAGA	AACCTCAGGG	AAACCAAAAC	TATATGAACA	
1801	TTCTCCGGTT	CCTGGCCCTG	GGCGACAAGA	AGCTGAGTCC	CTTTAACATC	
1851	ACTCAGCCCT	TCACGCACCT	CTTCTGGTTT	GGGGATCTTA	ACTACCGTGT	
1901	GGATCTGCCT	ACCTGGCAGG	CAGAAACCAT	CATCCAAAAA	ATCAAGCAGC	
1951	AGCAGTACGC	AGACCTCCTG	TCCCACGACC	AGCTGCTCAC	AGAGAGGAGG	
2001	GAGCAGAAGG	TCTTCTTACA	CTTCGAGGAG	GAAGAAATCA	CGTTTGCCCC	
2051	AACCTACCGT	TTTGAGAGAC	TGACTCGGGA	CAATACGCC	TACACCAAGC	
2101	AGAAAGCGAC	AGGGATGAAG	TACAACTTGC	CTTCTGGTGG	TGACCGAGTC	
2151	CTCTGGAAGT	CTTATCCCCC	GGTGACAGTG	GTGTGTCAGT	CTTATGGCAG	
2201	TACCAGCGAC	ATCATGACGA	GTGACCAACG	CCCTGTCTTT	GCCACATTTG	
2251	AGGCAGGAGT	CACCTCCACG	TTTGTCTCCA	AGAACGGTCC	CGGGACTGTT	
2301	GACAGCCAAAG	GACAGATTGA	GTTTCTCAGG	TGCTATGCCA	CATTGAAGAC	
2351	CAAGTCCACG	ACCAAAATCT	ACCTGGAGTT	CCACTCGAGC	TGCTTGGAGA	
2401	GTTTGTGTCAA	GAGTCAGGAA	GGAGAAAATG	AAGAAGGAAG	TGAGGGGGAG	
2451	CTGGTGGTGA	AGTTTGGTGA	GACTCTTCCA	AAGCTGAAGC	CCATTATCTC	
2501	TCACCCTGAG	TACCTGCTAG	ACCAGCACAT	CCTCATCAGC	ATCAAGTCTT	
2551	CTGACAGCGA	CGAATCCTAT	GGCGAGGGCT	GCATTGCCCT	TGGGTAGAG	
2601	GCCACAGAAA	CGCAGCTGCC	CATCTACACG	CCTCTCACC	ACCATGGGGA	
2651	GTTGACAGGC	CACCTCCAGG	GGGAGATCAA	GCTGCAGACC	TCTCAGGGCA	
2701	AGACGAGGGA	GAAGCTCTAT	GACTTGTGTA	AGACGGAGCG	TGATGAATCC	
2751	AGTGGGCCAA	AGACCCTGAA	GAGCCTCACC	AGCCACGACC	CCATGAAGCA	
2801	GTGGGAGATC	ACTAGCAGGC	CCCTCCGTG	CAGTGGCTCC	AGCATCACTG	
2851	AAATCATCAA	CCCCAACTAC	ATGGGAGTGG	GGCCCTTTGG	GCCACCAATG	
2901	CCCTGTCACG	TGAAGCAGAC	CTGTCCCCCT	GACCAGCAGC	CCACAGCCTG	
2951	GAGCTACGAC	CAGCCGCCCA	AGGACTCCCC	GCTGGGGCCC	TGCAGGGGAG	
3001	AAAGTCTCTC	GACACCTCCC	GGCCAGCCGC	CCATATCACC	CAAGAAAGTTT	

15/27

FIGURE 10 CONT'D

```

3051 TTACCCCTCAA CAGCAAACCG GGGTCTCCCT CCCAGGACAC AGGAGTCAAG
3101 GCCCAGTGAC CTGGGGAAGA ACGCAGGGGA CACGCTGCCT CAGGAGGACC
3151 TGCCCGTGAC GAAGCCCGAG ATGTTTGAGA ACCCCCTGTA TGGGTCCCTG
3201 AGTTCCCTTCC CTAAGCCTGC TCCCAGGAAG GACCAGGAAT CCCCCAAAAT
3251 GCCGCGGAAG GAACCCCGCG CCTGCCCGGA ACCCGGCATC TTGTGCCCCA
3301 GCATCGTGCT CACCAAAGCC CAGGAGGCTG ATCGCGGCGA GGGGCCCGGC
3351 AAGCAGGTGC CCGCGCCCCG GCTGCGCTCC TTCACGTGCT CATCTCTGCG
3401 CGAGGCGCAGG GCGGCCGCGG GGGACAAGAG CCAAGGGAAG CCCAAGACCC
3451 CGGTACGCTC CCAGGCCCCG GTGCCGCCA AGAGGCCCAT CAAGCCTTCC
3501 AGATCGGAAA TCAACCAGCA GACCCCGCCC ACCCGGACGC CGCGGCCGCC
3551 GCTGCCAGTC AAGAGCCCGG CGGTGCTGCA CCTCCAGCAC TCCAAGGGCC
3601 GCGACTACCG CGACAACACC GAGCTCCCGC ATCACGGCAA GCACCGGCCG
3651 GAGGAGGGGC CACCAGGGCC TCTAGGCAGG ACTGCCATGC AGTGAAGCCC
3701 TCAGTGAGCT GCCACTGAGT CCGGAGCCCA GAGGAACGGC GTGAAGCCAC
3751 TCGAGCCTCT CCGGGGACCT CCTGCTGCGT CCTCCTGCCC AGCTTCCTAT
3801 GCAAGGCTTT GTGTTTTAG GAAAGGCTT AGCTTCTGTG TGGCCACAG
3851 AGTTTCTGCG CTGTGAGGCT TAGCACCAAG TGCTGAGGCT GGAAGAAAAA
3901 CGCACACCAAG ACGGGCAACA AACAGTCTGG GTCCCCAGCT CGCTCTTGGT
3951 ACTTGGGACC CCAGTGCTC GTTGAGGGCG CCATTCTGAA GAAAGGAAT
4001 GCAGCGCCGA TTTGAGGGTG GAGATATAGA TAATAATAAT ATTAATAATA
4051 ATAATGGCCA CATGGATCGA ACACTCATGA TGTGCCAAGT GCTGTGCTAA
4101 GTGCTTTAGC AACATTGCTC ATATCAGGAT GACCTCGAGA GCTGAGGCTC
4151 TAGCCACCTA AAACACGTGC CCAAACCCAC CAGTTTAAAA CGTGTGTGT
4201 TCGGAGGGGT GAAAGCATTA AGAAGCCCAG TCCCCTCCTG GAGTGAGACA
4251 AGGGCTCGGC CTTAAGGAGC TGAAGAGTCT GGGTAGCTTG TTTAGGGTAC
4301 AAGAAGCCTG TTCTGTCCAG CTTCAAGTAC ACAAGCTGCT TTAGCTAAAG
4351 TCCCGCGGGT TCCGGCATGG CTAGGCTGAG AGCAGGGATC TACCTGGCTT
4401 CTCAGTTCTT TGGTTGGAAG GAGCAGGAAA TCAGCTCCTA TTCTCCAGTG
4451 GAGAGATCTG CCCTCAGCTT GGGCTAGAGA TGCCAAGGCC TGTGCCAGGT
4501 TCCCCTGCGC CTCTCCAGG TGGGCAGCCA TCACCAGCCA CAGTTAAGCC
4551 AAGCCCCCA ACATGTATTC CATCGTGCTG GTAGAAGAGT CTTTGTGTT
4601 GCTCCCGAAA GCGGTGCTCT CCAGCCTGGC TGCCAGGGAG GGTGGGCCCTC
4651 TTGTTCCAG GCTCTTGAAA TAGTGAGCC TTTCTTCCT ATCTCTGTGG
4701 CTTTCAGCTC TGCTTCCTTG GTTATTAGGA GAATAGATGG GTGATGTCTT
4751 TCCTTATGTT GCTTTTTCAA CATAGCAGAA TTAATGTAGG GAGCTAAATC
4801 CAGTGGTGTG TGTGAATGCA GAAGGGAATG CACCCACAT TCCCATGATG
4851 GAAGTCTGCG TAACCAATAA ATTGTGCTT TCTTAAAAAT TCGCGGCCGC
4901 GTCGACGTG ACCCGGCCGC GAATTC

```

STOP CODON

5' UNTRANSLATED REGION (3695-4925)

16/27  
**FIGURE 11**

**hSHIP Amino Acid Sequence**

1 MVPWNHGNI TRSKAEELLC RTGKDGSPV RASESIFRAY ALCVLYRNCV  
51 YTYRILPNED DKFTVQASEG VSMRFFTKLD QLIEFYKKEK MGLVTHLQYP  
101 VPLEEDTGD DPEEDTESVV SPPELPPRNI PLTASSCEAK EVPFSENENPR  
151 ATETSRPSLS ETLFQRLQSM DTSGLPEEHL KAIQDVLSTQ LAQDSEFVKY  
201 GSSSLPHLKK LTTLLCKELY CEVIRTLPSL ESLQRLFDQQ LSPGLRPRPQ  
251 VPGEANPINM VSKLSQLTSL LSSIEDKVKA LLHEGPESPH RPSLIPFVTF  
301 EVKAESLGIP QFMQLKVDVE SGKLIKKSK DGSEDKFYSH KKILQLIKSQ  
351 KFLNKLVLV ETEKEKILRK EYVFADSKR EGFCQLLOQM KKKHSEKPEP  
401 DMITIFIGTW NMGNAPPPKK ITSWPLSKGQ GKTRDDSDY IPHDIYVIGT  
451 QEDPLSEKEW LEILKHSLQE ITSVTPKIVA INTLWNIRIV VLAKPEHENR  
501 ISHICTDNVK TGIANTLGK GAVGVSPFN GTSLGFVNSH LTSGSEKLR  
551 RNQNYMNLK PLALGDKLS PFNITHRPTH LFWPGDLNVR VDLPTWEAET  
601 IIQIKKQQY ADLLSHDQLL TERREQVFL HFEEEEITFA PTYRFERLTR  
651 DKYAYTRQKA TGMKYNLPSW CDRVLWKSYP LVHVVCQSYG STSDIMTSDH  
701 SPVPATFEAG VTSQFVSKNG PGTVD SQGQI EPLRCYATLK TKSQTKFYLE  
751 FHSSCLESFV KSQEGENEKG SEGELVVKFG ETLPLKPII SDPEYLLDQH  
801 ILISIKSSDS DESYGEGCIA LRLEATETQL PIYTPLTHHG ELTGHPQGEI  
851 KLQTSQKTR EKLYDFVKE RDSSCPKTL KSLTSHDPMK QWEVTSRAPP  
901 CSGSSITEII NPNYMGVGPF GPPMPLHVQ TLSPDQOPTA WSYDQPPKDS  
951 PLGPCRGESP PTPPGQPPIS PKKFLPSTAN RGLPPRTQES RPSDLGKNAG  
1001 DTLPEEDLPL TKPEMFENPL YGSLSSFPKP APRKDQESPK MPRKEPPPCP  
1051 EPGILSPSIV LTKAQEADRG EGPQKQVPAP RLRSFTCSSS AEGRAAGGDK  
1101 SQGKPKTPVS SQAPVPAKRP IKPSRSEINQ QTPPTPTPRP PLPVKSPAVL  
1151 HLQHSKGRDY RDNTELPFHG KHRPEEGPPG PLGR TAMQ

17/27  
**FIGURE 12**

(Peptide) PASTA of: hshipcom.pep from: 1 to: 1188 April 3, 1996 13:17

TRANSLATE of: hshipcom.con check: 8429 from: 129 to: 3693.  
generated symbols 1 to: 1188.

TO: 145com.pep Sequences: 1 Symbols: 1,303 Word Size: 2  
Scoring matrix: GenRunData:fastapep.cmp  
Variable pamfactor used  
Gap creation penalty: 12.0 Gap extension penalty: 4.0

The best scores are:

initl initn opt..

/gcg/users/patty/145com.pep TRANSLATE of: 145com.con che...4283 4937 5189

hshipcom.pep

/gcg/users/patty/145com.pep

TRANSLATE of: 145com.con check: 4805 from: 130 to: 4040  
generated symbols 1 to: 1303.

SCORES Initl: 4283 Initn: 4937 Opt: 5189  
87.2% identity in 1194 aa overlap

	10	20	30	40	50
hshipc	MVPCWNHGNITRSKAEELLCRTGKDGSSFLVRASESIFRAYALCVLYRNCVYTYRILP				
145com	MPAMVPGWNHGNITRSKAEELLSRAGKDGSSFLVRASESIPRACALCVLFRNCVYTYRILP				
	10	20	30	40	50
	60	70	80	90	100
hshipc	NEDDKFTVQASEGVSMRFFTKLDQLIEFYKKENMGLVTHLQYPVPLEEEDTGDDPEEDTE				
145com	NEDDKFTVQASEGVPMRPFPTKLDQLIDFYKKENMGLVTHLQYPVPLEEEDAIDEAEDTE				
	70	80	90	100	110
	120	130	140	150	160
hshipc	SVVSPPPELPPRNIPLTASSCEAKEVPFSNENPRATETSRPSSLSETLPQRLQSMDSGLPE				
145com	SVMSPPPELPPRNIPMSAGPSRAKDLPLATENPRAPEVTRLSSLSETLPQRLQSMDSGLPE				
	130	140	150	160	170
	180	190	200	210	220
hshipc	EHLKAIQDYLLSTQLAQDSEFVKTGSSSLPHLKKLTTLCKELYGEVIRTLPSSLQRLP				
145com	EHLKAIQDYLLSTQLLLDSDFLKTGSSNLPHLKKLMSLLCKELHGEVIRTLPSSLQRLP				
	190	200	210	220	230
	240	250	260	270	280
hshipc	DQQLSPGLRPRPQVPGEANPINMVSKLSQLTSLSSIEDKVKALLHEGPESPHRPSLIYP				
145com	DQQLSPGLRPRPQVPGEASPTWAKLSQLTSLSSIEDKVKSLHEGSESTNRRLSIYP				
	250	260	270	280	290
	300	310	320	330	340
hshipc	VTFEVKAESLGIPQIMQLKVDVESGKLIHKSKDGSSEDKFYSHKKILQLIKSQKFLNKL				

**FIGURE 12 CONT'D**

[illegible]



20/27  
**FIGURE 13**

(Nucleotide) FASTA of: hshipcom.con from: 20 to: 4896 April 3, 1996 13:08

TO: 145com.con Sequences: 1 Symbols: 4,040 Word Size: 6  
 Scoring matrix: GenRunData:fastadna.cmp  
 Constant pamfactor used  
 Gap creation penalty: 12.0 Gap extension penalty: 4.0

The best scores are:

init1 initn opt..

/gcg/users/patty/145com.con

8658 10037 10667

hshipcom.con

/gcg/users/patty/145com.con

SCORES Init1: 8658 Initn: 10037 Opt: 10667  
 81.6% identity in 4019 bp overlap

	20	30	40	50
hshipc	CCCAAGAGGCAACGGGCGGCAGGTTGCAG--TGG			
145com	CCCTGGTAGGAGCAGCAGAGGCAATTTCTGAGAGGCAACAGGCGGCAGGTTCTCAGCCTAG			
	10	20	30	40
	50	60		
hshipc	60	70	80	90
	AGGGGGCCTCCGCTC-CCCTCGGTGGTGTGTGGGTCTGGGGGTGCTGCCGGCCAGCCG			
145com	70	80	90	100
	AGAGGGCCCTGAACTACTTTGCTGGAGTGTCCGTCCTGGGAGTGGCTGCTGACCCAGTCC			
	110	120		
hshipc	120	130	140	150
	AGGAGGGCCACGCCCACCATGGTCCCTGCTGGAACCATGGCAACATCACCCGCTCCAAG			
145com	130	140	150	160
	AGGAGACCCATGGCTGCCATGGTCCCTGGGTGGAACCATGGCAACATCACCCGCTCCAAG			
	170	180		
hshipc	180	190	200	210
	GCGGAGGAGCTGCTTTGCAGGACAGGCAAGGACGGGAGCTTCCTCGTGCGTGCCAGCGAG			
145com	190	200	210	220
	GCAGAGGAGCTACTTTCCAGAGCCGGCAAGGACGGGAGCTTCCTTGTGCGTGCCAGCGAG			
	230	240		
hshipc	240	250	260	270
	TCCATCTTCCGGGCATACGCGCTCTGCGTGCTGTATCGGAATTGCGTTTATACTTACAGA			
145com	250	260	270	280
	TCCATCCCCCGGCGCTGCGCACTCTGCGTGCTGTTCCGGAATTGTGTTTACACTTACAGG			
	290	300		
hshipc	300	310	320	330
	ATTCTGCCCAATGAAGATGATAAATTCAGGCTGTTTCAGGCATCCGAAGGCGTCTCCATGAGG			
145com	310	320	330	340
	ATTCTGCCCAATGAGGACGATAAATTCAGGCTGTTTCAGGCATCCGAAGGTGTCCCATGAGG			
	350	360		

21/27

FIGURE 13 CONT'D

	360	370	380	390	400	410
hshipc	TTCTTCACCAAGCTGGACCAGCTCATCGAGTTT	TACAAGAAGGAAAACATGGGGCTGGTG				
145com	TTCTTCACGAAGCTGGACCAGCTCATCGACTTT	TACAAGAAGGAAAACATGGGGCTGGTG				
	370	380	390	400	410	420
	420	430	440	450	460	470
hshipc	ACCCATCTGCAATACCTGTGCCGCTGGAGGAAGAGGACACAGCGACGACCTGAGGAG					
145com	ACCCACCTGCAGTACCCCGTGCCCTGGAGGAGGAGGATGCTATTGATGAGGCTGAGGAG					
	430	440	450	460	470	480
	480	490	500	510	520	530
hshipc	GACACAGAAAGTGTCTGTCTCCACCCGAGCTGCCCCAAGAAACATCCCGCTGACTGCC					
145com	GACACTGAAAGTGTCTATGTCAACACCTGAGCTGCCTCCAGAAACATTCCTATGTCTGCC					
	490	500	510	520	530	540
	540	550	560	570	580	590
hshipc	AGCTCCTGTGAGGCCAAGGAGGTTCCCTTTTCAAACGAGAATCCCCGAGCGACCGAGACC					
145com	GGGCCCAGCGAGGCCAAGGACCTTCCTCTTGCAACAGAGAACCCCCGAGCCCCCTGAGGTC					
	550	560	570	580	590	600
	600	610	620	630	640	650
hshipc	AGCCGGCCGAGCCTCTCCGAGACATTGTTCCAGCGACTGCAAAGCATGGACACCACTGGG					
145com	ACCCGGCTGAGTCTCTCCGAGACACTGTTTCAGCGTCTACAGAGCATGGATACCACTGGG					
	610	620	630	640	650	660
	660	670	680	690	700	710
hshipc	CTTCCAGAAGAGCATCTTAAGGCCATCCAAGATTATTTAAGCACTCAGCTCGCCAGGAC					
145com	CTTCCCGAGGAGCACCTGAAAGCCATCCAGGATTATCTGAGCACTCAGCTCCTCTGGAT					
	670	680	690	700	710	720
	720	730	740	750	760	770
hshipc	TCTGAATTGTGAAGACAGGGTCCAGCAGTCTTCCTCACCTGAAGAACTGACCACACTG					
145com	TCCGACTTTTGAAGAACGGGCTCCAGCAACTCCCTCACCTGAAGAAGCTGATGTCACTG					
	730	740	750	760	770	780
	780	790	800	810	820	830
hshipc	CTCTGCAAGGAGCTCTATGGAGAAGTCATCCGGACCCTCCCATCCCTGGAGTCTCTGCAG					
145com	CTCTGCAAGGAGCTCCATGGGGAAGTCATCAGGACTCTGCCATCCCTGGAGTCTCTGCAG					
	790	800	810	820	830	840
	840	850	860	870	880	890
hshipc	AGGTTATTGACCAGCAGCTCTCCCGGGCCTCCGTCCACGTCCTCAGGTTCTGGTGAG					
145com	AGGTTGTTTGACCAACAGCTCTCCCGAGGCCTTCGCCACGACCTCAGGTGCCCGGAGAG					
	850	860	870	880	890	900
	900	910	920	930	940	950
hshipc	GCCAATCCCATCAACATGGTGTCCAAGCTCAGCCAACCTGACAAGCCTGTTGTATCCATT					
145com	GCCAGTCCCATCACCATGGTTGCCAACTCAGCCAATTGACAAGTCTGTCTCTCCATT					
	910	920	930	940	950	960

22/27

**FIGURE 13 CONT'D**

	960	970	980	990	1000	1010
hshipc	GAAGACAAGGTCAAGGCCTTGCTGCACGAGGGTCCTGAGTCTCCGCACCGCCCTCCCTT					
145com	GAAGATAAGGTCAAGTCCTTGCTGCACGAGGGCTCAGAATCTACCAACAGGCGTTCCTT					
	970	980	990	1000	1010	1020
	1020	1030	1040	1050	1060	1070
hshipc	ATCCCTCCAGTCACCTTTGAGGTGAAGGCAGAGTCTCTGGGGATTCTCAGAAAATGCAG					
145com	ATCCCTCCGGTCACCTTTGAGGTGAAGTCAGAGTCCCTGGGCATTCTCAGAAAATGCAT					
	1030	1040	1050	1060	1070	1080
	1080	1090	1100	1110	1120	1130
hshipc	CTCAAAGTCGACGTTGAGTCTGGGAAACTGATCATTAAAGAAGTCCAAGGATGGTTCTGAG					
145com	CTCAAAGTGGACGTTGAGTCTGGGAAACTGATCGTTAAAGAAGTCCAAGGATGGTTCTGAG					
	1090	1100	1110	1120	1130	1140
	1140	1150	1160	1170	1180	1190
hshipc	GACAAGTTCTACAGCCACAAGAAAATCCTGCAGCTCATTAAAGTCACAGAAATTTCTGAAT					
145com	GACAAGTTCTACAGCCACAAAAAATCCTGCAGCTCATTAAAGTCCCAGAAGTTTCTAATC					
	1150	1160	1170	1180	1190	1200
	1200	1210	1220	1230	1240	1250
hshipc	AAGTTGGTGATCTTGTTGGTGAACAGAGAAGGAGAAGATCTGCGGAAGGAATATGTTTTT					
145com	AAGTTGGTGATTTTGGTGGAGACGGAGAAGGAGAAAATCCTGAGGAAGGAATATGTTTTT					
	1210	1220	1230	1240	1250	1260
	1260	1270	1280	1290	1300	1310
hshipc	GCTGACTCCAAAAAGAGAGAAGGCTTCTGCCAGCTCCTGCAGCAGATGAAGAACAGCAC					
145com	GCTGACTCTAAGAAAAGAGAAGGCTTCTGTCAACTCCTGCAGCAGATGAAGAACAGCAT					
	1270	1280	1290	1300	1310	1320
	1320	1330	1340	1350	1360	1370
hshipc	TCAGAGCAGCCGGAGCCCGACATGATCACCATCTTCATCGGCACCTGGAACATGGGTAAC					
145com	TCCGAGCAGCCAGAGCCTGACATGATCACCATCTTCATTGGCACTTGGAAACATGGGTAAT					
	1330	1340	1350	1360	1370	1380
	1380	1390	1400	1410	1420	1430
hshipc	GCCCCCCTCCCAAGAAGATCACGTCTGGTTTCTCTCCAAGGGGCAGGGAAAGACGCGG					
145com	GCACCCCTCCCAAGAAGATCACGTCTGGTTTCTCTCCAAGGGGCAGGGAAAGACACCG					
	1390	1400	1410	1420	1430	1440
	1440	1450	1460	1470	1480	1490
hshipc	GACGACTCTGCGGACTACATCCCCCATGACATTTACGTGATCGGCACCCAAGAGGACCCC					
145com	GACGACTCTGCTGACTACATCCCCCATGACATCTATGTGATTGGCAOCCAGGAGGATCCC					
	1450	1460	1470	1480	1490	1500
	1500	1510	1520	1530	1540	1550
hshipc	CTGAGTGAGAAGGAOTGGCTGGAGATCCTCAAACTCCCTGCAAGAAATCACCAGTGTG					
145com	CTTGGAGAGAAGGAGTGGCTGGAGCTACTCAGGCCTCCCTGCAAGAAATCACCAGCATG					
	1510	1520	1530	1540	1550	1560

23/27

FIGURE 13 CONT'D

	1560	1570	1580	1590	1600	1610
hshipc	ACTTTTAAAACAGTCGCCATCCACACGCTCTGGAACATCCGCATCGTGGTGCTGGCCAAG					
145com	ACATTTAAAACAGTTGCCATCCACACCTCTGGAACATTGCGATAGTGGTGCTTGCCAAG					
	1570	1580	1590	1600	1610	1620
	CCTGAGCAGGAGAACCGGATCAGCCACATCTGTACTGACAACGTGAAGACAGGCATTGCA					
hshipc	CCTGAGCAGGAGAACCGGATCAGCCACATCTGTACTGACAACGTGAAGACAGGCATTGCA					
145com	CCAGAGCATGAGAATCGGATCAGCCATATCTGCACTGACAACGTGAAGACAGGCATTGCA					
	1630	1640	1650	1660	1670	1680
	AACACACTGGGGAACAAGGGAGCCGTGGGGGTGTCGTTTCATGTTCAATGGAACCTCCTTA					
hshipc	AACACACTGGGGAACAAGGGAGCCGTGGGGGTGTCGTTTCATGTTCAATGGAACCTCCTTA					
145com	AACACCTTGGGAAACAAGGGAGCAGTGGGAGTGTCTTCATGTTCAATGGAACCTCCTTG					
	1690	1700	1710	1720	1730	1740
	GGGTTTCGTCAACAGCCACTTGACTTCAGGAAGTGAAAAGAACTCAGGCGAAACCAAAAC					
hshipc	GGGTTTCGTCAACAGCCACTTGACTTCAGGAAGTGAAAAGAACTCAGGCGAAACCAAAAC					
145com	GGGTTTCGTCAACAGCCACTTGACTTCAGGAAGTGAAAAGAACTCAGGCGAAACCAAAAC					
	1750	1760	1770	1780	1790	1800
	TATATGAACATTCTCCGGTTCCTGGCCCTGGGCGACAAGAAGCTGAGTCCCTTTAACATC					
hshipc	TATATGAACATTCTCCGGTTCCTGGCCCTGGGCGACAAGAAGCTGAGTCCCTTTAACATC					
145com	TATATGAACATTCTCCGGTTCCTGGCCCTGGGCGACAAGAAGCTAAGCCCATTTAACATC					
	1810	1820	1830	1840	1850	1860
	ACTCACCCTTCACGCACCTCTTCTGGTTTGGGGATCTTAACTACCGTGTGGATCTGCGCT					
hshipc	ACTCACCCTTCACGCACCTCTTCTGGTTTGGGGATCTTAACTACCGTGTGGATCTGCGCT					
145com	ACCCACCCTTCACCCACCTCTTCTGGCTTGGGGATCTCAACTACCGCGTGGAGCTGCCC					
	1870	1880	1890	1900	1910	1920
	ACCTGGGAGGCAGAAACCATCATCCAAAAATCAAGCAGCAGCAGTACGCAGACCTCCTG					
hshipc	ACCTGGGAGGCAGAAACCATCATCCAAAAATCAAGCAGCAGCAGTACGCAGACCTCCTG					
145com	ACTTGGGAGGCAGAGGCCATCATCCAGAAGATCAAGCAACAGCAGTATTGAGACCTTCTG					
	1930	1940	1950	1960	1970	1980
	TCCCACGACCAGCTGCTCACAGAGAGGAGGAGCAGAAAGTCTTCTTACACTTCGAGGAG					
hshipc	TCCCACGACCAGCTGCTCACAGAGAGGAGGAGCAGAAAGTCTTCTTACACTTCGAGGAG					
145com	GCCCACGACCAACTGCTCCTGGAGAGGAAGGACCAGAAAGTCTTCTTGCACCTTTGAGGAG					
	1990	2000	2010	2020	2030	2040
	GAAGAAATCACGTTTGCCCCAACCTACCGTTTGGAGAGTGAAGTCTTCTTACACTTCGAGGAG					
hshipc	GAAGAAATCACGTTTGCCCCAACCTACCGTTTGGAGAGTGAAGTCTTCTTACACTTCGAGGAG					
145com	GAAGAGATCACCTTCGCCCCAACCTATCGATTGAAAGACTGACCCGGGACAAAGTATGCA					
	2050	2060	2070	2080	2090	2100
	TACACCAAGCAGAAAGCGACAGGGATGAAGTACAACCTTGCCCTTCTGGTGTGACCGAGTC					
hshipc	TACACCAAGCAGAAAGCGACAGGGATGAAGTACAACCTTGCCCTTCTGGTGTGACCGAGTC					
145com	TACACGAAGCAGAAAGCAACAGGGATGAAGTACAACCTTGCCCTTCTGGTGTGACCGAGTC					
	2110	2120	2130	2140	2150	2160

24/27

**FIGURE 13 CONT'D**

2160 2170 2180 2190 2200 2210  
hshipc CTCTGGAAGTCTTATCCCCTGGTGCACGTGGTGTGTCTCAGTCTTATGGCAGTACCAGCGAC  
|||||  
145com CTCTGGAAGTCTTATCCCCTGGTGCATGTGGTCTGTCTCAGTCTTATGGCAGTACCAGTGC  
2170 2180 2190 2200 2210 2220  
2220 2230 2240 2250 2260 2270  
hshipc ATCATGACGAGTGACCACAGCCCTGTCTTTGCCACATTTGAGGCAGGAGTCACTTCCAG  
|||||  
145com ATCATGACGAGTGACCACAGCCCTGTCTTTGCCACGTTTGAAGCAGGAGTCACTCTCAA  
2230 2240 2250 2260 2270 2280  
2280 2290 2300 2310 2320 2330  
hshipc TTTGTCTCCAAGAACGGTCCCGGGACTGTGTGACAGCCAAGGACAGATTGAGTTTCTCAGG  
|||||  
145com TTCGTCTCCAAGAATGGTCTCTGGCAGTGTAGATAGCCAAGGGCAGATCGAGTTTCTTGCA  
2290 2300 2310 2320 2330 2340  
2340 2350 2360 2370 2380 2390  
hshipc TGCTATGCCACATTGAAGACCAAGTCCCAGACCAAATTTCTACCTGGAGTTCCACTCGAGC  
|||||  
145com TGCTACGCCACACTGAAGACCAAGTCCCAGACTAAGTTCTACTTGGAGTTCCACTCAAGC  
2350 2360 2370 2380 2390 2400  
2400 2410 2420 2430 2440 2450  
hshipc TGCTTGGAGAGTTTGTCAAGAGTCAGGAAGGAGAGAAATGAAGAAGGAAGTGAGGGGGAG  
|||||  
145com TGCTTAGAGAGTTTGTCAAGAGTCAGGAAGGAGAGAGAAATGAAGAGGGAGTGAGGAGAG  
2410 2420 2430 2440 2450 2460  
2460 2470 2480 2490 2500 2510  
hshipc CTGGTGGTGAAGTTTGGTGAGACTCTTCCAAAGCTGAAGCCCATTATCTCTGACCCCTGAG  
|||||  
145com CTGGTGGTACGGTTTGGGAGAGACTCTTCCCAAGCTAAAGCCGATTATCTCTGACCCCGAG  
2470 2480 2490 2500 2510 2520  
2520 2530 2540 2550 2560 2570  
hshipc TACCTGCTAGACCAGCACATCCTCATCAGCATCAAGTCTCTGACAGCGACGAATCCTAT  
|||||  
145com TACTTACTGGACCAGCATATCCTGATCAGCATTAATCCTCTGACAGTGACGAGTCTAT  
2530 2540 2550 2560 2570 2580  
2580 2590 2600 2610 2620 2630  
hshipc GGGGAGGGCTGCATTGCCCTTCGGTTAGAGGCCACAGAAACGCAGCTGCCCATCTACAGC  
|||||  
145com GGTGAAGGGCTGCATTGCCCTTCGGTTAGAGGCCACAGAGGCTCAGCATCTATCTACAGC  
2590 2600 2610 2620 2630 2640  
2640 2650 2660 2670 2680 2690  
hshipc CCTCTCACCACCATGGGGAGTTGACAGGCCACTTCCAGGGGGAGATCAAGCTGCAGACC  
|||||  
145com CCTCTCACCACCATGGGGAGATGACTGGCCACTTCCAGGGGAGAGATTAAGCTGCAGACC  
2650 2660 2670 2680 2690 2700  
2700 2710 2720 2730 2740 2750  
hshipc TCTCAGGGCAAGACGAGGGAGAAGCTCTATGACTTTGTGAAGACGAGCGTGATGAATCC  
|||||  
145com TCCCAGGGCAAGATGAGGGAGAAGCTCTATGACTTTGTGAAGACAGAGCGGGATGAATCC  
2710 2720 2730 2740 2750 2760

**FIGURE 13 CONT'D**

2760 2770 2780 2790 2800 2810

hshipc AGTGGGCCAAGACCCTGAAGAGCCTCACCAGCCACGACCCCATGAAGCAGTGGGAAGTC  
||||| || ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

145com AGTGGGAATGAAATGCTTGAAGAACCTCACCAGCCATGACCTTATGAGGCAATGGGAGCCT  
2770 2780 2790 2800 2810 2820

2820 2830 2840 2850 2860 2870

hshipc ACTAGCAGGGCCCCCTCCGTGCAGTGGCTCCAGCATCACTGAAATCATCAACCCCAACTAC  
|| ||||||| ||||| || ||||||| ||||||| ||||||| ||||||| |||||||

145com TCTGCCAGGGTCCCTGCATGTGTGTCTCCAGCCTCAATGAGATGATCAATCCAAACTAC  
2830 2840 2850 2860 2870 2880

2880 2890 2900 2910 2920 2930

hshipc ATGGGAGTGGGGCCCTTTGGGGCCACCAATGCCCCCTGCACGTGAAGCAGACCTTGTCCCCT  
|| || ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

145com ATTGGTATGGGGCCTTTTGG-----ACAGCCCCTGCATGGGAAATCAACCCCTGTCCCCA  
2890 2900 2910 2920 2930

2940 2950 2960 2970 2980 2990

hshipc GACCAGCAGCCACAGCCTGGAGCTACGACCAGCCGCCCCAAGGACTCCCCGCTGGGGCCC  
|| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

145com GATCAGCAACTCAGAGCTTGGAGTTATGACCAGCTACCCAAAGACTCCTCCCTGGGGCCT  
2940 2950 2960 2970 2980 2990

3000 3010 3020 3030 3040 3050

hshipc TGCAGGGGAGAAAGTCTCTCGACACCTCCCGGCCAGCCGCCCATATCACCCAAGAAGTTT  
|| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

145com GGGAGGGGGGAGGGTCTCCAACCCCTCCCTCCCAACCACCTCTGTGCGCCAAGAAGTTT  
3000 3010 3020 3030 3040 3050

3060 3070 3080 3090 3100 3110

hshipc TTACCCTCAACAGCAAACCGGGGTCTCCCTCCCAGGACACAGGAGTCAAGGCCCAGTGAC  
|| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

145com TCATCTTCCACAACCAACCGAGGTCCCTGCCCCAGGGTGCAAGAGGCCAAGACCTGGGGAT  
3060 3070 3080 3090 3100 3110

3120 3130 3140 3150 3160 3170

hshipc CTGGGGAAGAACGCAGGGGACAGCTGCCTCAGGAGGACCTGCCGCTGACGAAGCCCGAG  
|| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

145com CTGGGGAAG-----GTGGAAGCTCTGCTCCAGGAGGACCTGCTGCTGACGAAGCCCGAG  
3120 3130 3140 3150 3160

3180 3190 3200 3210 3220 3230

hshipc ATGTTTGAGAACCCCTGTATGGGTCCCTGAGTTCCTTCCCTAAGCCTGCTCCAGGAAG  
|| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

145com ATGTTTGAGAACCCACTGTATGGATCCGTGAGTTCCTTCCCTAAGCTGGTGCCAGGAAA  
3170 3180 3190 3200 3210 3220

3240 3250 3260 3270 3280 3290

hshipc GACCAGGAATCCCCAAAATGCCGCGGAAGGAACCCCGCCCTGCCCGGAACCCGGCATC  
|| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

145com GAGCAGGAGTCTCCCAAGATGCTGCGGAAGGAGCCCCCGCCCTGTCCAGACCCAGGAATC  
3230 3240 3250 3260 3270 3280

3300 3310 3320 3330 3340 3350

hshipc TTGTCGCCCAGCATCGTGCTACCCAAAGCCCAGGAGGCTGATCGCGGCGAGGGGCCCCGGC  
|| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

145com TCATCACCAGCATCGTGCTCCCAAGCCCAAGAGGCTGGAGAGTGTCAAGGGGACAAGC  
3290 3300 3310 3320 3330 3340

**FIGURE 13 CONT'D**

[illegible]

27/27

FIGURE 13 CONT'D

```

      3920      3930      3940      3950      3960      3970
hshipc GCAACAAACAGTCTG-GGTCCCCAG--CTCGCTCTTGTACTTGGGACCCCACTGCCTCG
      ||||| ||||| ||||| ||||| ||||| |||||
145com ACGCCATACAGACAGCAGACAGCGGCACTGGGTCTCAGAACTT-GGATTCTGGGCCTTC
      3940      3950      3960      3970      3980      3990

      3980      3990      4000      4010      4020      4030
hshipc TTGAGGGCGCCATTCTGAAGAAAGGAACTGCAGCGCCGATTGAGGGTGGAGATATAGAT
      || ||||| ||||| ||||| ||||| |||||
145com TTCCAGTCGCCGTTTAAAGAAAGGAACTAACGGAGCTGCTCATCCGA
      4000      4010      4020      4030      4040

```

! CPU time used:

! Database scan: 0:00:00.8

! Post-scan processing: 0:00:01.4

! Total CPU time: 0:00:02.4

! Output File: b.

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**